(51) International Patent Classification⁷: C07K 14/705

(21) International Application Number: PCT/US01/21174

(22) International Filing Date: 2 July 2001 (02.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/215,495	30 June 2000 (30.06.2000)	US
60/215,483	30 June 2000 (30.06.2000)	US
60/216,117	6 July 2000 (06.07.2000)	US
60/220,589	25 July 2000 (25.07.2000)	US
60/245,294	2 November 2000 (02.11.2000)	US
60/260,729	10 January 2001 (10.01.2001)	US
60/260,851	10 January 2001 (10.01.2001)	US
60/271,673	26 February 2001 (26.02.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US	60/260,851 (CIP)
Filed on	10 January 2001 (10.01.2001)
US	60/215,495 (CIP)
Filed on	30 June 2000 (30.06.2000)
US	60/271,673 (CIP)
Filed on	26 February 2001 (26.02.2001)
US	60/215,483 (CIP)
Filed on	30 June 2000 (30.06.2000)
US	60/260,729 (CIP)
Filed on	10 January 2001 (10.01.2001)
US	60/216,117 (CIP)
Filed on	6 July 2000 (06.07.2000)
US	60/220,589 (CIP)
Filed on	25 July 2000 (25.07.2000)
US	60/245,294 (CIP)
Filed on	2 November 2000 (02.11.2000)

(71) Applicant (for all designated States except US): CURA-GEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): SPADERNA, Steven, K. [US/US]; 261 Deerfiled Drive, Berlin, CN 06037 (US). PADIGARU, Muralidhara [IN/US]; 98 Montoya Circle, Branford, CT 06405 (US). SPYTEK, Kimberly, A. [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). CASMAN, Stacie [US/US]; 155 Knoll-wood Drive, Wallingford, CT 06492 (US). RASTELLI, Luca [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). MAYANKAR, Uriel [IN/US]; 35 Averill Place, Branford, CT 06405 (US). TCHERNEV, Velizar [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The inventio further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

BACKGROUND

The invention generally relates to novel GPCR1, GPCR2, GPCR3, GPCR4, GPCR5, GPCR6, GPCR7, GPCR8, GPCR9 and GPCR10 nucleic acids and polypeptides encoded

5   therefrom. More specifically, the invention relates to nucleic acids encoding novel polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides. More particularly,

10   the invention relates to nucleic acids encoding novel G-protein coupled receptor (GPCR) polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding

15   novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as GPCRX, or GPCR1, GPCR2, GPCR3, GPCR4, GPCR5, GPCR6, GPCR7, GPCR8, GPCR9 and GPCR10 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "GPCRX" nucleic acid or polypeptide sequences.

20   In one aspect, the invention provides an isolated GPCRX nucleic acid molecule encoding a GPCRX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27. In some embodiments, the GPCRX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes

25   a protein-coding sequence of a GPCRX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a GPCRX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10,

1

12, 14, 16, 18, 21, 23, 25 and 28. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a GPCRX nucleic acid (*e.g.*, SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27) or a complement of said oligonucleotide.

Also included in the invention are substantially purified GPCRX polypeptides (SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28). In certain embodiments, the GPCRX polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human GPCRX polypeptide.

The invention also features antibodies that immunoselectively bind to GPCRX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a GPCRX nucleic acid, a GPCRX polypeptide, or an antibody specific for a GPCRX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a GPCRX nucleic acid, under conditions allowing for expression of the GPCRX polypeptide encoded by the DNA. If desired, the GPCRX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a GPCRX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the GPCRX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a GPCRX.

Also included in the invention is a method of detecting the presence of a GPCRX nucleic acid molecule in a sample by contacting the sample with a GPCRX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a GPCRX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a GPCRX polypeptide by contacting a cell sample that includes the GPCRX polypeptide with a compound that binds to the GPCRX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic

5   acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting

10   disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, or other disorders related to cell signal processing and metabolic pathway modulation. The therapeutic can be, *e.g.*, a GPCRX nucleic acid, a GPCRX polypeptide, or a GPCRX-specific

15   antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability Disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, Retinal diseases including those involving

20   photoreception, Cell growth rate disorders; cell shape disorders, feeding disorders; control of feeding; potential obesity due to over-eating; potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer),

25   anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease; multiple sclerosis; Albright Hereditary Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, allergies, benign prostatic hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation. Dentatorubro-pallidoluysian

30   atrophy (DRPLA) Hypophosphatemic rickets, autosomal dominant (2) Acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders of the like.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and

3

antagonist compounds. For example, a cDNA encoding GPCRX may be useful in gene
therapy, and GPCRX may be useful when administered to a subject in need thereof. By way
of nonlimiting example, the compositions of the present invention will have efficacy for
treatment of patients suffering from bacterial, fungal, protozoal and viral infections

5   (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to
Neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia,
asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention,
osteoporosis, Crohn's disease; multiple sclerosis; and Treatment of Albright Hereditary
Ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, allergies, benign prostatic

10   hypertrophy, and psychotic and neurological disorders, including anxiety, schizophrenia,
manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as
Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders.
       The invention further includes a method for screening for a modulator of disorders or
syndromes including, e.g., diabetes, metabolic disturbances associated with obesity, the

15   metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases,
metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated
cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder,
immune disorders, and hematopoietic disorders or other disorders related to cell signal
processing and metabolic pathway modulation. The method includes contacting a test

20   compound with a GPCRX polypeptide and determining if the test compound binds to said
GPCRX polypeptide. Binding of the test compound to the GPCRX polypeptide indicates the
test compound is a modulator of activity, or of latency or predisposition to the aforementioned
disorders or syndromes. .
       Also within the scope of the invention is a method for screening for a modulator of

25   activity, or of latency or predisposition to an disorders or syndromes including, e.g., diabetes,
metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting
disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious
disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders,
Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders or

30   other disorders related to cell signal processing and metabolic pathway modulation by
administering a test compound to a test animal at increased risk for the aforementioned
disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a
GPCRX nucleic acid. Expression or activity of GPCRX polypeptide is then measured in the
test animal, as is expression or activity of the protein in a control animal which recombinantly-

expresses GPCRX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of GPCRX polypeptide in both the test animal and the control animal is compared. A change in the activity of GPCRX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or

5   syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a GPCRX polypeptide, a GPCRX nucleic acid, or both, in a subject (e.g., a human subject). The method includes measuring the amount of the GPCRX polypeptide in a test sample from the subject and

10   comparing the amount of the polypeptide in the test sample to the amount of the GPCRX polypeptide present in a control sample. An alteration in the level of the GPCRX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, e.g., diabetes, metabolic disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders

15   associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

20   In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a GPCRX polypeptide, a GPCRX nucleic acid, or a GPCRX-specific antibody to a subject (e.g., a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, e.g., diabetes, metabolic

25   disturbances associated with obesity, the metabolic syndrome X, anorexia, wasting disorders associated with chronic diseases, metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders.

In yet another aspect, the invention can be used in a method to identity the cellular

30   receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention

belongs. Although methods and materials similar or equivalent to those described herein can

be used in the practice or testing of the present invention, suitable methods and materials are

described below. All publications, patent applications, patents, and other references

mentioned herein are incorporated by reference in their entirety. In the case of conflict, the

5   present specification, including definitions, will control. In addition, the materials, methods,

and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following

detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

10   The invention is based, in part, upon the discovery of novel nucleic acid sequences that

encode novel polypeptides. The novel nucleic acids and their encoded polypeptides are

referred to individually as GPCR1, GPCR2, GPCR3, GPCR4, GPCR5, GPCR6, GPCR7,

GPCR8, GPCR9 and GPCR10. The nucleic acids, and their encoded polypeptides, are

collectively designated herein as "GPCRX".

15   The novel GPCRX nucleic acids of the invention include the nucleic acids whose

sequences are provided in Tables 1A, 1C, 2A, 2C, 3A, 4A, 5A, 5C, 6A, 6C, 7A, 8A, 9A, 9B

and 10A, inclusive, or a fragment, derivative, analog or homolog thereof. The novel GPCRX

proteins of the invention include the protein fragments whose sequences are provided in

Tables 1B, 1D, 2B, 2D, 3B, 4B, 5B, 5D, 6B, 7B, 8B, 9B and 10B, inclusive. The individual

20   GPCRX nucleic acids and proteins are described below. Within the scope of this invention is

a method of using these nucleic acids and peptides in the treatment or prevention of a disorder

related to cell signaling or metabolic pathway modulation.

The GPCRX proteins of the invention have a high homology to the 7tm_1 domain

(PFam Acc. No. pfam00001). The 7tm_1 domain from the 7 transmembrane receptor family,

25   which includes a number of different proteins, including, for example, serotonin receptors,

dopamine receptors, histamine receptors, andrenergic receptors, cannabinoid receptors,

angiotensin II receptors, chemokine receptors, opioid receptors, G-protein coupled receptor

(GPCR) proteins, olfactory receptors (OR), and the like. Some proteins and the Protein Data

Base Ids/gene indexes include, for example: rhodopsin (129209); 5-hydroxytryptamine

30   receptors; (112821, 8488960, 112805, 231454, 1168221, 398971, 112806); G protein-coupled

receptors (119130, 543823, 1730143, 132206, 137159, 6136153, 416926, 1169881, 136882,

134079); gustatory receptors (544463, 462208); c-x-c chemokine receptors (416718, 128999,

416802, 548703, 1352335); opsins (129193, 129197, 129203); and olfactory receptor-like proteins (129091, 1171893, 400672, 548417).

Because of the close homology among the members of the GPCRX family, proteins that are homologous to any one member of the family are also largely homologous to the other

5   members, except where the sequences are different as shown below.

The similarity information for the GPCRX proteins and nucleic acids disclosed herein suggest that GPCR1-GPCR10 may have important structural and/or physiological functions characteristic of the Olfactory Receptor family and the GPCR family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic

10   applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in

15   gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration in vitro and in vivo (vi) biological defense weapon.

G-Protein Coupled Receptor proteins (GPCRs) have been identified as a large family of G protein-coupled receptors in a number of species. These receptors share a seven transmembrane domain structure with many neurotransmitter and hormone receptors, and are

20   likely to underlie the recognition and G-protein-mediated transduction of various signals. Human GPCR generally do not contain introns and belong to four different gene subfamilies, displaying great sequence variability. These genes are dominantly expressed in olfactory epithelium. See, *e.g.*, Ben-Arie et al., *Hum. Mol. Genet.* 1994 3:229-235; and, Online Mendelian Inheritance in Man (OMIM) entry # 164342 (http://www.ncbi.nlm.nih.gov/entrez/

25   dispomim.cgi?).

The olfactory receptor (OR) gene family constitutes one of the largest GPCR multigene families and is distributed among many chromosomal sites in the human genome. See Rouquier et al., *Hum. Mol. Genet.* 7(9):1337-45 (1998); Malnic et al., *Cell* 96:713-23 (1999). Olfactory receptors constitute the largest family among G protein-coupled receptors,

30   with up to 1000 members expected. See Vanderhaeghen et al., *Genomics* 39(3):239-46 (1997); Xie et al., *Mamm. Genome* 11(12):1070-78 (2000); Issel-Tarver et al., *Proc. Natl. Acad. Sci. USA* 93(20):10897-902 (1996). The recognition of odorants by olfactory receptors is the first stage in odor discrimination. See Krautwurst et al., *Cell* 95(7):917-26 (1998); Buck et al., *Cell* 65(1):175-87 (1991). Many ORs share some characteristic sequence motifs and

have a central variable region corresponding to a putative ligand binding site. See Issel-Tarver et al., *Proc. Natl. Acad. Sci. USA* 93:10897-902 (1996).

Other examples of seven membrane spanninG proteins that are related to GPCRs are chemoreceptors. See Thomas et al., Gene 178(1-2):1-5 (1996). Chemoreceptors have been identified in taste, olfactory, and male reproductive tissues. See *id.*; Walensky et al., *J. Biol. Chem.* 273(16):9378-87 (1998); Parmentier et al., *Nature* 355(6359):453-55 (1992); Asai et al., *Biochem. Biophys. Res. Commun.* 221(2):240-47 (1996).

The GPCRX nucleic acids and proteins are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further below. For example, a cDNA encoding the GPCR (or olfactory-receptor) like protein may be useful in gene therapy, and the receptor -like protein may be useful when administered to a subject in need thereof. The nucleic acids and proteins of the invention are also useful in potential therapeutic applications used in the treatment of developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, retinal diseases including those involving photoreception, cell growth rate disorders, cell shape disorders, feeding disorders, potential obesity due to over-eating, potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, allergies, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, Crohn's disease, multiple sclerosis, Albright hereditary ostoeodystrophy, angina pectoris, myocardial infarction, ulcers, benign prostatic hypertrophy, psychotic and neurological disorders (including anxiety, schizophrenia, manic depression, delirium, dementia, and severe mental retardation), dentatorubro-pallidoluysian atrophy (DRPLA), hypophosphatemic rickets, autosomal dominant (2) acrocallosal syndrome and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. Other GPCR-related diseases and disorders are contemplated.

The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding the GPCR-like protein may be useful in gene therapy, and the GPCR-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will

have efficacy for treatment of patients suffering from developmental diseases, MHCII and III diseases (immune diseases), taste and scent detectability disorders, Burkitt's lymphoma, corticoneurogenic disease, signal transduction pathway disorders, retinal diseases including those involving photoreception, cell growth rate disorders, cell shape disorders, feeding

5 disorders, potential obesity due to over-eating, potential disorders due to starvation (lack of appetite), noninsulin-dependent diabetes mellitus (NIDDM1), bacterial, fungal, protozoal and viral infections (particularly infections caused by HIV-1 or HIV-2), pain, cancer (including but not limited to neoplasm; adenocarcinoma; lymphoma; prostate cancer; uterus cancer), anorexia, bulimia, asthma, allergies, Parkinson's disease, acute heart failure, hypotension,

10 hypertension, urinary retention, osteoporosis, Crohn's disease, multiple sclerosis, Albright hereditary osteoedystrophy, angina pectoris, myocardial infarction, ulcers, benign prostatic hypertrophy, psychotic and neurological disorders (including anxiety, schizophrenia, manic depression, delirium, dementia, and severe mental retardation), dentatorubro-pallidoluysian atrophy (DRPLA), hypophosphatemic rickets, autosomal dominant (2) acrocallosal syndrome

15 and dyskinesias, such as Huntington's disease or Gilles de la Tourette syndrome and/or other pathologies and disorders. The novel nucleic acid encoding GPCR-like protein, and the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind

20 immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

GPCR1

A GPCR-like protein of the invention, referred to herein as GPCR1, is an Olfactory Receptor ("OR")-like protein. The novel GPCR1 nucleic acid sequences were identified on

25 chromosome 11 as described in Example 1. Some members of the Olfactory Receptor-Like Protein Family end up localized at the cell surface, where they exhibit activity. Therefore it is likely that these novel GPCR1 proteins are available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

The following genomic clone was identified as having regions with high homology to

30 the homolog. Genomic clone >acc:AP001804 HTG Homo sapiens chromosome 11 clone RP11-164A10 map 11q, WORKING DRAFT SEQUENCE, in unordered pieces - Homo sapiens, 165058 bp (DNA) was analyzed by Genscan and Grail software to identify exons and putative coding sequences.

Two alternative novel GPCR1 nucleic acids and encoded polypeptides are provided, namely GPCR1a and GPCR1b.

GPCR1a

In one embodiment, a GPCR1 variant is the novel GPCR1a (alternatively referred to
5   herein as CG54326_02), which includes the 977 nucleotide sequence (SEQ ID NO:1) shown
in Table 1A. A GPCR1a ORF begins with a Kozak consensus ATG initiation codon at
nucleotides 24-26 and ends with a TGA codon at nucleotides 957-959. The DNA sequence
and protein sequence for a GPCR1a gene or one of its splice forms was obtained solely by
exon linking. Putative untranslated regions upstream from the initiation codon and
10   downstream from the termination codon are underlined in Table 1A, and the start and stop
codons are in bold letters.

Table 1A. GPCR1 Nucleotide Sequence (SEQ ID NO:1)
TTACACATAATACCTTAAAAGACATGGCTACTTCAAACCATTCTTCAGGGGCTGAGTTTATCCTGGC
AGGCTTGACACAACGCCCAGAACTTCAACTGCCACTCTTCCTCCTGTTCCTTGGAATATATGTGGTC
ACAGTGGTGGGGAACCTGGGCATGATCTTCTTAATTGCTCTCAGTTCTCAACTTTACCCTCCAGTGT
ATTATTTTCTCAGTCATTTGTCTTTCATTGATCTCTGCTACTCCTCTGTCATTACCCCTAAGATGCT
GGTGAACTTTGTTCCAGAGGAGAACATTATCTCCTTTCTGGAATGCATTACTCAACTTTATTTCTTC
CTTATTTTTGTAATTGCAGAAGGCTACCTTCTGACAGCCATGGAATGTGACCGTTATGTTGCTATCT
GTCGCCCACTGCTTTACAATATTGTCATGTCCCACAGGGTCTGTTCCATAATGATGGCTGTGGTATA
CTCACTGGGTTTTCTGTGGGCCACAGTCCATACTACCCGCATGTCAGTGTTGTCATTCTGTAGGTCT
CATACGGTCAGTCATTATTTTTGTGATATTCTCCCCTTATTGACTCTGTCTTGCTCCAGCACCCACA
TCAATGAGATTCTGCTGTTCATTATTGGAGGAGTTAATACCTTAGCAACTACACTGGCGGTCCTTAT
CTCTTATGCTTTCATTTTCTCTAGTATCCTTGGTATTCATTCCACTGAGGGGCAATCCAAAGCCTTT
GGCACTTGTAGCTCCCATCTCTTGGCTGTGGGCATCTTTTTTGGGTCTATAACATTCATGTATTTCA
AGCCCCCTTCCAGCACTACTATGGAAAAAGAGAAGGTGTCTTCTGTGTTCTACATCACAATAATCCC
CATGCTGAATCCTCTAATCTATAGCCTGAGGAACAAGGATGTGAAAAATGCACTGAAGAAGATGACT
AGGGGAAGGCAGTCATCCTGA<u>CAAAGAGGGTTCTCATTG</u>

The cDNA coding for the GPCR1a sequence was cloned by the polymerase chain
15   reaction (PCR). Primers were designed based on *in silico* predictions of the full length or some
portion (one or more exons) of the cDNA/protein sequence of the invention. *In silico*
prediction was based on sequences available in CuraGen's proprietary sequence databases or
in the public human sequence databases, and provided either the full length DNA sequence, or
some portion thereof. The DNA sequence and protein sequence for a novel GPCR1 gene were
20   obtained by exon linking and are reported here as GPCR1a. These primers and methods used to
amplify GPCR1a cDNA are described in the Examples.

The GPCR1a polypeptide (SEQ ID NO:2) encoded by SEQ ID NO:1 is 311 aa in
length, has a molecular weight of 34795.35 Daltons, and is presented using the one-letter
amino acid code in Table 1B. The Psort profile for both GPCR1a and GPCR1b predicts that

these sequences have a signal peptide and are likely to be localized at the plasma membrane
with a certainty of 0.600. In alternative embodiments, a GPCR1 polypeptide is located to the
Golgi body with a certainty of 0.400, the endoplasmic reticulum (membrane) with a certainty
of 0.300, or a microbody (peroxisome) with a certainty of 0.300. The Signal P predicts a

5   likely cleavage site for a GPCR1 peptide is between positions 41 and 42, *i.e.*, at the dash in the
sequence VVG-NL.

Table 1B. GPCR1a protein sequence (SEQ ID NO:2)
MATSNHSSGAEFILAGLTQRPELQLPLFLLFLGIYVVTVVGNLGMIFLIALSSQLYPPVYYFLSHLS FIDLCYSSVITPKMLVNFVPEENIISFLECITQLYFFLIFVIAEGYLLTAMECDRYVAICRPLLYNI VMSHRVCSIMMAVVYSLGFLWATVHTTRMSVLSFCRSHTVSHYFCDILPLLTLSCSSTHINEILLFI IGGVNTLATTLAVLISYAFIFSSILGIHSTEGQSKAFGTCSSHLLAVGIFFGSITFMYFKPPSSTTM EKEKVSSVFYITIIPMLNPLIYSLRNKDVKNALKKMTRGRQSS

GPCR1b

10   In an alternative embodiment, a GPCR1 variant is the novel GPCR1b (alternatively
referred to herein as AP001804_A), which includes the 936 nucleotide sequence (SEQ ID
NO:3) shown in Table 1C. The GPCR1b ORF begins with a Kozak consensus ATG initiation
codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 934-936, which are in
bold letters in Table 1C.

15

Table 1C. GPCR1b Nucleotide Sequence (SEQ ID NO:3)
ATGGCTACTTCAAACCATTCTTCAGGGGCTGAGTTTATCCTGGCAGGCTTGACACAACGCCCAGAAC TTCAACTGCCACTCTTCCTCCTGTTCCTTGGAATATATGTGGTCACAGTGGTGGGGAACCTGGGCAT GATCTTCTTAATTGCTCTCAGTTCTCAACTTTACCCTCCAGTGTATTATTTTCTCAGTCATTTGTCT TTCATTGATCTCTGCTACTCCTCTGTCATTACCCCTAAGATGCTGGTGAACTTTGTTCCAGAGGAGA ACATTATCTCCTTTCTGGAATGCATTACTCAACTTTATTTCTTCCTTATTTTTGTAATTGCAGAAGG CTACCTTCTGACAGCCATGGAATATGACCGTTATGTTGCTATCTGTCGCCCACTGCTTTACAATATT GTCATGTCCCACAGGGTCTGTTCCATAATGATGGCTGTGGTATACTCACTGGGTTTTCTGTGGGCCA CAGTCCATACTACCCGCATGTCAGTGTTGTCATTCTGTAGGTCTCATACGGTCAGTCATTATTTTTG TGATATTCTCCCCTTATTGACTCTGTCTTGCTCCAGCACCCACATCAATGAGATTCTGCTGTTCATT ATTGGAGGAGTTAATACCTTAGCAACTACACTGGCGGTCCTTATCTCTTATGCTTTCATTTTCTCTA GTATCCTTGGTATTCATTCCACTGAGGGGCAATCCAAAGCCTTTGGCACTTGTAGCTCCCATCTCTT GGCTGTGGGCATCTTTTTTGGGTCTATAACATTCATGTATTTCAAGCCCCCTTCCAGCACTACTATG GAAAAAGAGAAGGTGTCTTCTGTGTTCTACATCACAATAATCCCCATGCTGAATCCTCTAATCTATA GCCTGAGGAACAAGGATGTGAAAAATGCACTGAAGAAGATGACTAGGGGAAGGCAGTCATCC**TGA**

The GPCR1b protein (SEQ ID NO:4) encoded by SEQ ID NO:3 is 311 amino acid in
length, has a molecular weight of 34855.38 Daltons, and is presented using the one-letter code
in Table 1D. As with GPCR1a, the most likely cleavage site for a GPCR1b peptide is between

20   amino acids 41 and 42, *i.e.*, at the dash in the sequence VVG-NL, based on the SignalP result.

Table 1D. GPCR1b protein sequence (SEQ ID NO:4)
MATSNHSSGAEFILAGLTQRPELQLPLFLLFLGIYVVTVVGNLGMIFLIALSSQLYPPVYYFLSHLS FIDLCYSSVITPKMLVNFVPEENIISFLECITQLYFFLIFVIAEGYLLTAMEYDRYVAICRPLLYNI VMSHRVCSIMMAVVYSLGFLWATVHTTRMSVLSFCRSHTVSHYFCDILPLLTLSCSSTHINEILLFI IGGVNTLATTLAVLISYAFIFSSILGIHSTEGQSKAFGTCSSHLLAVGIFFGSITFMYFKPPSSTTM EKEKVSSVFYITIIPMLNPLIYSLRNKDVKNALKKMTRGRQSS

GPCR1 Clones

Unless specifically addressed as GPCR1a or GPCR1b, any reference to GPCR1 is
assumed to encompass all variants. Residue differences between any GPCRX variant
5   sequences herein are written to show the residue in the "a" variant, the residue position with
respect to the "a" variant, and the residue in the "b" variant. For example, the GPCR1 nucleic
acid sequences differ at the following position: G382A. The GPCR1 polypeptides differ only
at one residue, namely C120Y. The homologies shown above are shared by GPCR1b insofar
as GPCR1a and 1b are homologous as shown in Table 1E and Table 1G. GPCRX residues in
10   all following sequence alignments that differ between the individual GPCRX variants are
marked with the (o) symbol above the variant residue in all alignments herein.

The amino acid sequence of GPCR1 had high homology to other proteins as shown in
Table 1E.

Table 1E. BLASTX results for GPCR1			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY90875 Human G protein-coupled receptor GTAR11-1	+1	1092	9.6e-110
>patp:AAY90877 Human G protein-coupled receptor GTAR11-3	+1	979	9.0e-98

15   In a search of sequence databases, it was found, for example, that the GPCR1 nucleic
acid sequence has 657 of 932 bases (70%) identical to a gb:GENBANK-ID:RNOLP4
|acc:X80671.1 mRNA from *Rattus norvegicus* (*R. norvegicus* olp4 mRNA). The full GPCR1
amino acid sequence was found to have 209 of 305 amino acid residues (68%) identical to,
and 253 of 305 amino acid residues (82%) similar to, the 309 amino acid residue
20   ptnr:SPTREMBL-ACC:Q63395 protein from Rattus norvegicus (Rat) (OLFACTORY
RECEPTOR). In all BLAST herein, the "E-value" or "Expect" value is a numeric indication
of the probability that the aligned sequences could have achieved their similarity to the
BLAST query sequence by chance alone, within the database that was searched. Additional
BLAST results are shown in Table 1F.

25

| Table 1F. BLAST results for GPCR1					
Gene Index/ Identifier	**Protein/ Organism**	**Length (aa)**	**Identity (%)**	**Positives (%)**	**Expect**
gi|14423803|sp|Q9GZ M6; gi|12002782|gb|AAG4 3386.1| (AF162668); gi|12002784|gb|AAG4 3387.1|AF162669_1 (AF162669)	olfactory receptor-like protein JCG2 [Homo sapiens]	311	301/311 (96%)	301/311 (96%)	e-145
gi|11692559|gb|AAG3 9876.1|AF282291_1 (AF282291)	gi|11692559|gb|AAG 39876.1|AF282291_1 (AF282291)	308	241/307 (78%)	264/307 (85%)	e-117
gi|11692555|gb|AAG3 9874.1|AF282289_1 (AF282289)	odorant receptor K40 [Mus musculus]	308	209/308 (67%)	247/308 (79%)	e-101
gi|1083741|pir||S51 356; gi|517366|emb|CAA56 697.1| (X80671)	olfactory receptor [Rattus norvegicus]	309	202/305 (66%)	246/305 (80%)	8e-98
gi|11692557|gb|AAG3 9875.1|AF282290_1 (AF282290)	odorant receptor K41 [Mus musculus]	308	195/305 (63%)	241/305 (78%)	5e-92

A multiple sequence alignment is given in Table 1G, with the GPCR1 protein of the invention being shown on line 1, in a ClustalW analysis comparing GPCR1 with related protein sequences disclosed in Table 1F. The residue that differs between GPCR1a and

5 GPCR1b is marked with the (o) symbol.

Table 1G. Information for the ClustalW proteins:

1. >GPCR1; SEQ ID NO:4
2. >gi\|14423803\|sp\|Q9GZM6\|O8D2_Human Olfactory Receptor 8D2 (OR-Like Protein JCG2); SEQ ID NO:29
3. >gi\|11692559\|gb\|AAG39876.1\|AF282291_1 odorant receptor K42 [Mus musculus]; SEQ ID NO:30
10 4. >gi\|11692555\|gb\|AAG39874.1\|AF282289_1 odorant receptor K40 [Mus musculus]; SEQ ID NO:31
5. >gi\|1083741\|pir\|\|S51356 olfactory receptor - rat; SEQ ID NO:32
6. >gi\|11692557\|gb\|AAG39875.1\|AF282290_1 odorant receptor K41 [Mus musculus]; SEQ ID NO:33

```
                                                 o
GPCR1          QLYFFLIFVLAEGYLLTAMEYDRYVAICRPLLYNIVMSHRVCSIMMAVVY
gi|14423803|   QLYFFLIFVLAEGYLLTAMEYDRYVAICRPLLYNIVMSHRVCSIMMAVVY
gi|11692559|   QLYFFLLLAISEGYLLTAMAYDRYVAICSPLLYNTVMSHKVCSIMMAVVY
gi|11692555|   QLFFFVIFVVAEGYLLTAMAYDRYVAICRPLLYNVIMSSRLCSLLVLVAF
gi|1083741|    QFFFFAIFVVTEGYLLTVMAYDRYVAICRPLLYNVIMSSRICSLLVLVAF
gi|11692557|   QFFFFAVFVVTEGYLLTVMAYDHYVAICRPLLYNVMMSSKHCLLLVLVAF

                    160       170       180       190       200
               ....|....|....|....|....|....|....|....|....|....|
GPCR1          SLGFLWATVHTDRMSVLSFCRSHTVSHYFCDILPLLTLSCSSTHINEILL
gi|14423803|   SLGFLWATVHTDRMSVLSFCRSHTVSHYFCDILPLLTLSCSSTHINEILL
gi|11692559|   SLGFFGATVHTTRMTMLSFCGSHIIRHYFCDILPLLTLSCSSTHINEVLL
gi|11692555|   ILGFVSALAHTSAMMNLSFCKSHVISHYFCDVLPLLNLSCSDIKLNELLL
gi|1083741|    SLGLFSAVVHTSAMMNLSFCKSYIISHYFCDALPLLKLACSNTHLNELLL
gi|11692557|   TLGLFSAVVHTSAMMSLNFCKTYIISHYFCDALPLLKLSCSNTHLNELLL

                    210       220       230       240       250
               ....|....|....|....|....|....|....|....|....|....|
GPCR1          FIIGGVNTLATTLAVLISYAFIFSSILGIHSTEGQSKAFGTCSSHLLAVG
gi|14423803|   FIIGGVNTLATTLAVLISYAFIFSSILGIHSTEGQSKAFGTCSSHLLAVG
gi|11692559|   FIIGGVNTLAPTLAVIISYAFILTSILRIRSNEGRSKAFGTCSSHIMAVG
gi|11692555|   FIIAGFNTLVPTLAVAISYVFIFCSILHIKSSKGRSKAFGTCSSHLMAVG
gi|1083741|    FIIGGLNTLVPTLAVAISYVFIFCSILRIRSSEGRSKAFGTCSSHLMAVG
gi|11692557|   FIIGGINTLVPTLAVAISYVFIFCSIRHIKSSKSRSKAFGTCSSHLMAVG

                    260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|
GPCR1          IFFGSITFMYFKPPSSTTMEKEKVSSVFYITIIPMLNPLIYSLRNKDVKN
gi|14423803|   IFFGSITFMYFKPPSSTTMEKEKVSSVFYITIIPMLNPLIYSLRNKDVKN
gi|11692559|   IFFGSITFMYFKPPSSNNMEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKT
gi|11692555|   IFFGSITFMYFKPPSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|1083741|    IFFGSITFMYLKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|11692557|   IFFGSITFMYLKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK

                    310
               ....|....|..
GPCR1          ALKKMTRGRQSS
gi|14423803|   ALKKMTRGRQSS
gi|11692559|   ALKKMVGRRQLS
gi|11692555|   ALGKCLAGR---
gi|1083741|    ALGRFSVRS---
gi|11692557|   ALGRFSVRR---
```

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). The results indicate that the GPCR1 protein contains the following protein domain (as defined by Interpro): domain name 7tm_1 7 transmembrane receptor (rhodopsin family). DOMAIN results for GPCR1 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections.

As discussed below, all GPCRX proteins of the invention contain significant homology to the 7tm_1 domain. This indicates that the GPCRX sequence has properties similar to those of other proteins known to contain this 7tm_1 domain and similar to the properties of these domains. The 254 amino acid domain termed 7tm_1 (SEQ ID NO:34), a seven

5   transmembrane receptor (rhodopsin family), is shown in Table 1H.


Table 1H. 7tm_1, 7 transmembrane receptor domain

gnl|Pfam|pfam00001, 7tm_1, 7 transmembrane receptor (rhodopsin family). (SEQ ID NO:34)
GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDWVFGDALCKLVGALFVVNGYASILLLTAISIDRYL AIVHPLRYRRIRTPRRAKVLILLVWVLALLLSLPPLLFSWLRTVEEGNTTVCLIDFPEESVKRSYVLLSTLVGFVLPLLVILVC YTRILRTLRKRARSQRSLKRRSSSERKAAKMLLVVVVVVFVLCWLPYHIVLLLDSLCLLSIWRVLPTALLITLWLAYVNSCLNPI IY

10   Table 1I lists the domain description from DOMAIN analysis results against GPCR1. This indicates that the GPCR1a sequence has properties similar to those of other proteins known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID NO:34). For Table 1I and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading and "strong" semi-conserved residues are indicated by

15   grey shading. The "strong" group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

The DOMAIN results are listed in Table 1I with the statistics and domain description. An alignment of GPCR1 residues 41-290 (SEQ ID NOs:2 and 4) with the full 7tm_1 domain,

20   residues 1-254 (SEQ ID NO:34), are shown in Table 1I.

Table 1I. DOMAIN results for GPCR1

PSSMs producing significant alignments:	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	96.3	2e-21

```
                    10        20        30        40        50
               ....|....|....|....|....|....|....|....|....|....|
25
   GPCR1        GNLGMILLIAVSPLLHTPMYYFLSSLSFVDFCYSSVLTPKMLVNFLGKKN
   Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDW

                    60        70        80        90        100
               ....|....|....|....|....|....|....|....|....|....|
30
   GPCR1        TILYSECMVQLFRFVVFVVAEGYLLTAMAYDRYVAICSPLLYNAIMSSWV
   Pfam|pfam00001 VFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPRR

                    110       120       130       140       150
```

```
                ....|....|....|....|....|....|....|....|....|....|....|
GPCR1           CSLLVLAAFFLGFLSALTHTSAMMKLSFCKSHIINH----YFCDVLPLLN
Pfam|pfam00001  AKVLILLVWVLALLLSLPP---LLFSWLRTVEEGNT----TVCLIDF---

                   160       170       180       190       200
                ....|....|....|....|....|....|....|....|....|....|
GPCR1           LSCSNTH--LNELLLFIIAGFNTLVPTLAVAVSYAFII---------Y--
Pfam|pfam00001  -PEESVK--RSYVLLSTLVGF--VLPLLVILVCYTRIIRTLRKRARSQ-*

                   210       220       230       240       250
                ....|....|....|....|....|....|....|....|....|....|
GPCR1           --------------------------------------------------S
Pfam|pfam00001  --------------------------------------------------R

                   260       270       280       290       300
                ....|....|....|....|....|....|....|....|....|....|
GPCR1           ILHIRSSEGRSKAFGTCSSHLMAVVIFFGSITFMYF----KPPSS-----
Pfam|pfam00001  SLKRRSSSERKAAKMLLVVVVVVFVLCWLPYHIVLLLDSLCLLSIW-----

                   310       320
                ....|....|....|....|....|.
GPCR1           -NSLDQEKVSSVFYTTVIPMLNPLIY
Pfam|pfam00001  -RVLPTALLITLWLAYVNSCLNPILY
```

The nucleic acids and proteins of GPCR1 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, as described further herein.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR1 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR1 epitope is from about amino acids 10 to 20. In another embodiment, a GPCR1 epitope is from about amino acids 175 to 190. In specific embodiments, GPCR1 epitopes are from about amino acids 230 to 245, from about amino acids 258 to 273 and from about amino acids 290 to 311.

GPCR2

A second GPCR-like protein of the invention, referred to herein as GPCR2, is an Olfactory Receptor ("OR")-like protein. The novel GPCR2 nucleic acid sequences were identified on chromosome 11 as described in Example 1. Some members of the Olfactory Receptor-Like Protein Family end up localized at the cell surface, where they exhibit activity.

Therefore it is likely that these novel GPCR2 proteins are available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

The following genomic clone was identified as having regions with high homology to the homolog. Genomic clone >acc:AP001804 HTG Homo sapiens chromosome 11 clone

5   RP11-164A10 map 11q, WORKING DRAFT SEQUENCE, in unordered pieces - Homo sapiens, 165058 bp (DNA) was analyzed by Genscan and Grail software to identify exons and putative coding sequences.

Two alternative novel GPCR2 nucleic acids and encoded polypeptides are provided, namely GPCR2a and GPCR2b.

10   **GPCR2a**

In one embodiment, a GPCR2 variant is the novel GPCR2a (alternatively referred to herein as CG54335_02), which includes the 954 nucleotide sequence (SEQ ID NO:5) shown in Table 2A. A GPCR2a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 14-16 and ends with a TGA codon at nucleotides 938-940. Putative untranslated

15   regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 2A, and the start and stop codons are in bold letters.

Table 2A. GPCR2 Nucleotide Sequence (SEQ ID NO:5)
TGCCTAAAGAAGAATGACCATGGAAAATTATTCTATGGCAGCTCAGTTTGTCTTAGATGGTTTAACA CAGCAAGCAGAGCTCCAGCTGCCCCTCTTCCTCCTGTTCCTGGGAATCTATGTGGTCACAGTAGTGG GCAACCTGGGCATGATTCTCCTGATTGCAGTCAGCCCTCTACTTCACACCCCCATGTACTATTTCCT CAGCAGCTTGTCCTTCGTCGATTTCTGCTATTCCTCTGTCATTACTCCCAAAATGCTGGTGAACTTC CTAGGAAAGAAGAATACAATCCTTTACTCTGAGTGCATGGTCCAGCTCTTTTTCTTTGTGGTCTTTG TGGTGGCTGAGGGTTACCTCCTGACTGCCATGGCATATGATCGCTATGTTGCCATCTGGAGCCCACT GCTTTATAATGCGATCATGTCCTCATGGGTCTGCTCACTGCTAGTGCTGGCTGCCTTCTTCTTGGGC TTTCTCTCTGCCTTGACTCATACAAGTGCCATGATGAAACTGTCCTTTTGCAAATCCCACATTATCA ACCATTACTTCTGTGATGTTCTTCCCCTCCTCAATCTCTCCTGCTCCAACACACACCTCAATGAGCT TCTACTTTTTATCATTGCGGGGTTTAACACCTTGGTGCCCACCCTAGCTGTTGCTGTCTCCTATGCC TTCATCCTCTACAGCATCCTTCACATCCGCTCCTCAGAGGGCCGGTCCAAAGCTTTTGGAACATGCA GCTCTCATCTCATGGCTGTGGTGATCTTCTTTGGGTCCATTACCTTCATGTATTTCAAGCCCCCTTC AAGTAACTCCCTGGACCAGGAGAAGGTGTCCTCTGTGTTCTACACCACGGTGATCCCCATGCTGAAC CCTTTAATATACAGTCTGAGGAATAAGGATGTGAAGAAAGCATTAAGGAAGGTCTTAGTAGGAAAAT GAGTCCTGATTTGGGG

The sequence of GPCR2a was derived by laboratory cloning of cDNA fragments, by *in*

20   *silico* prediction of the sequence. The cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was based on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The cDNA coding for the GPCR2a sequence was cloned by the polymerase chain reaction (PCR). Primers were designed based on *in silico* predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. The DNA sequence and protein sequence for a novel Olfactory Receptor-like gene were obtained by

5   exon linking and are reported here as GPCR2a.These primers and methods used to amplify GPCR2 a cDNA are described in the Examples.

The GPCR2a polypeptide (SEQ ID NO:6) encoded by SEQ ID NO:5 is 308 aa in length, has a molecular weight of 34526.32 Daltons, and is presented using the one-letter amino acid code in Table 2B. The Psort profile for both GPCR2a and GPCR2b predicts that

10   these sequences have a signal peptide and are likely to be localized at the plasma membrane with a certainty of 0.600. In alternative embodiments, a GPCR2 polypeptide is located to the Golgi body with a certainty of 0.400, the endoplasmic reticulum (membrane) with a certainty of 0.300, or a microbody (peroxisome) with a certainty of 0.300. The Signal P predicts a likely cleavage site for a GPCR2 peptide is between positions 41 and 42, *i.e.*, at the dash in the

15   sequence VVG-NL.

Table 2B. GPCR2a protein sequence (SEQ ID NO:6)
MTMENYSMAAQFVLDGLTQQAELQLPLFLLFLGIYVVTVVGNLGMILLIAVSPLLHTPMYY FLSSLSFVDFCYSSVITPKMLVNFLGKKNTILYSECMVQLFFFVVFVVAEGYLLTAMAYDR YVAIWSPLLYNAIMSSWVCSLLVLAAFFLGFLSALTHTSAMMKLSFCKSHIINHYFCDVLP LLNLSCSNTHLNELLLFIIAGFNTLVPTLAVAVSYAFILYSILHIRSSEGRSKAFGTCSSH LMAVVIFFGSITFMYFKPPSSNSLDQEKVSSVFYTTVIPMLNPLIYSLRNKDVKKALRKVL VGK

GPCR2b

In an alternative embodiment, a GPCR2 variant is the novel GPCR2b (alternatively referred to herein as AP001804_B), which includes the 927 nucleotide sequence (SEQ ID

20   NO:7) shown in Table 2C. The GPCR2b ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 925-927, which are in bold letters in Table 2C.

Table 2C. GPCR2b Nucleotide Sequence (SEQ ID NO:7)
ATGACCATGGAAAATTATTCTATGGCAGCTCAGTTTGTCTTAGATGGTTTAACACAGCAAG CAGAGCTCCAGCTGCCCCTCTTCCTCCTGTTCCTGGGAATCTATGTGGTCACAGTAGTGGG CAACCTGGGCATGATTCTCCTGATTGCAGTCAGCCCTCTACTTCACACCCCCATGTACTAT TTCCTCAGCAGCTTGTCCTTCGTCGATTTCTGCTATTCCTCTGTCATTACTCCCAAAATGC TGGTGAACTTCCTAGGAAAGAAGAATACAATCCTTTACTCTGAGTGCATGGTCCAGCTCTT TTTCTTTGTGGTCTTTGTGGTGGCTGAGGGTTACCTCCTGACTGCCATGGCATATGATCGC TATGTTGCCATCTGTAGCCCACTGCTTTATAATGCGATCATGTCCTCATGGGTCTGCTCAC TGCTAGTGCTGGCTGCCTTCTTCTTGGGCTTTCTCTCTGCCTTGACTCATACAAGTGCCAT

```
GATGAAACTGTCCTTTTGCAAATCCCACATTATCAACCATTACTTCTGTGATGTTCTTCCC
CTCCTCAATCTCTCCTGCTCCAACACACACCTCAATGAGCTTCTACTTTTTATCATTGCGG
GGTTTAACACCTTGGTGCCCACCCTAGCTGTTGCTGTCTCCTATGCCTTCATCCTCTACAG
CATCCTTCACATCCGCTCCTCAGAGGGCCGGTCCAAAGCTTTTGGAACATGCAGCTCTCAT
CTCATGGCTGTGGTGATCTTCTTTGGGTCCATTACCTTCATGTATTTCAAGCCCCCTTCAA
GTAACTCCCTGGACCAGGAGAAGGTGTCCTCTGTGTTCTACACCACGGTGATCCCCATGCT
GAACCCTTTAATATACAGTCTGAGGAATAAGGATGTGAAGAAAGCATTAAGGAAGGTCTTA
GTAGGAAAATGA
```

The GPCR2b protein (SEQ ID NO:8) encoded by SEQ ID NO:7 is 308 amino acid in length, has a molecular weight of 34443.26 Daltons, and is presented using the one-letter code in Table 2D. As with GPCR2a, the most likely cleavage site for a GPCR2b peptide is between amino acids 41 and 42, i.e., at the dash in the sequence VVG-NL, based on the SignalP result.

Table 2D. GPCR2b protein sequence (SEQ ID NO:8)

```
MTMENYSMAAQFVLDGLTQQAELQLPLFLLFLGIYVVTVVGNLGMILLIAVSPLLHTPMYY
FLSSLSFVDFCYSSVITPKMLVNFLGKKNTILYSECMVQLFFFVVFVVAEGYLLTAMAYDR
YVAICSPLLYNAIMSSWVCSLLVLAAFFLGFLSALTHTSAMMKLSFCKSHIINHYFCDVLP
LLNLSCSNTHLNELLLFIIAGFNTLVPTLAVAVSYAFILYSILHIRSSEGRSKAFGTCSSH
LMAVVIFFGSITFMYFKPPSSNSLDQEKVSSVFYTTVIPMLNPLIYSLRNKDVKKALRKVL
VGK
```

GPCR2 Clones

Unless specifically addressed as GPCR2a or GPCR2b, any reference to GPCR2 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant, the residue position with respect to the "a" variant, and the residue in the "b" variant. For example, the GPCR2 nucleic acid sequences differ at the following position: G394T. The GPCR2 polypeptides differ only at one residue, namely W127C.

The amino acid sequence of GPCR2 had high homology to other proteins as shown in Table 2E.

Table 2E. BLASTX results for GPCR2			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY90875 Human G protein-coupled receptor GTAR11-1	+1	1484	2.8e-151
>patp:AAY90879 Human G protein-coupled receptor GTAR11-1	+1	1260	1.5e-127

In a search of sequence databases, it was found, for example, that the GPCR2 nucleic acid sequence of this invention has 770 of 922 bases (83%) identical to a *Rattus norvegicus* Olfactory Receptor-like protein mRNA (GENBANK-ID:RNOLP4|acc:X80671). The full amino acid sequence of the protein of the invention was found to have 247 of 302 amino acid

19

residues (81%) identical to, and 261 of 302 residues (86%) similar to, the 309 amino acid

residue Olfactory Receptor-like protein from *Rattus norvegicus* (SPTREMBL-ACC:Q63395).

In all BLAST herein, the "E-value" or "Expect" value is a numeric indication of the

probability that the aligned sequences could have achieved their similarity to the BLAST

5   query sequence by chance alone, within the database that was searched. In addition, it was

found, for example, that the GPCR2 nucleic acid sequence has 778 of 927 bases (83%)

identical to a gb:GENBANK-ID:AF282289| acc:AF282289.1 mRNA from Mus musculus

(Mus musculus odorant receptor K40 gene, complete cds). The full amino acid sequence of the

protein of the invention was found to have 259 of 304 amino acid residues (85%) identical to,

10   and 275 of 304 amino acid residues (90%) similar to, the 308 amino acid residue

ptnr:TREMBLNEW-ACC:AAG39874 protein from Mus musculus (Mouse) (ODORANT

RECEPTOR K40). Additional BLAST results are shown in Table 2F.

Table 2F. BLAST results for GPCR2					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11692555|gb|AAG3 9874.1|AF282289_1 (AF282289)	odorant receptor K40 [Mus musculus]	308	233/304 (76%)	246/304 (80%)	e-109
gi|1083741|pir||S51356 gi|517366|emb|CAA56 697.1| (X80671)	olfactory receptor [Rattus norvegicus]	309	223/302 (73%)	235/302 (76%)	e-102
gi|11692559|gb|AAG3 9876.1|AF282291_1 (AF282291)	odorant receptor K42 [Mus musculus]	308	207/301 (68%),	235/301 (77%)	e-100
gi|11692557|gb|AAG3 9875.1|AF282290_1 (AF282290)	odorant receptor K41 [Mus musculus]	308	219/306 (71%)	232/306 (75%)	e-100
gi|10644515|gb|AAG2 1322.1|AF271049_1 (AF271049)	odorant receptor [Mus musculus]	268	211/260 (81%)	220/260 (84%)	2e-95

A multiple sequence alignment is given in Table 2G, with the GPCR2 protein of the

15   invention being shown on line 1, in a ClustalW analysis comparing GPCR2 with related

protein sequences of Table 2F. The residue that differs between GPCR2a and GPCR2b is

marked with the (o) symbol.

Table 2G. Information for the ClustalW proteins:

20   1. >GPCR2; SEQ ID NO:5
2. >gi\|11692555\|gb\|AAG39874.1\|AF282289_1 odorant receptor K40 [Mus musculus]; SEQ ID NO:35
3. >gi\|1083741\|pir\|\|S51356 olfactory receptor - rat; SEQ ID NO:36
4. >gi\|11692559\|gb\|AAG39876.1\|AF282291_1 odorant receptor K42 [Mus musculus]; SEQ ID NO:37
5. >gi\|11692557\|gb\|AAG39875.1\|AF282290_1 odorant receptor K41 [Mus musculus]; SEQ ID NO:38

6. >gi|10644515|gb|AAG21322.1|AF271049_1 odorant receptor [Mus musculus]; SEQ ID NO:39


```
                    10        20        30        40        50
GPCR2          -MTMENYSMAAQFVLDGLTQQAELQLPLFILFLGIYVVTVVGNLGMILLI
gi|11692555|   -MGTGNHSVTVVFVLVGLTQQPELLLPLFILFLGIYVVTAVGNLGMILLI
gi|1083741|    MMGTGNHSAVVVFVLVGLTKQPELLLPLFFLFLVIYVITVVGNLGMILLI
gi|11692559|   ----MNHSVTDFILEGLTKRPELQLPLFILFLGIEVITVVGNLGMILLI
gi|11692557|   -MATGNHSAAVVFVLVGLTQQPELLLPLFILFLGIYVVTAVGNLGMILLI
gi|10644515|   ------------------------LPLFILFLGIYVVTAVGNLGMILLI

                    60        70        80        90        100
GPCR2          AVSPLLHTPMYYFLSSLSFVDFCYSSVITPKMLVNFLGKKNTILYSECMV
gi|11692555|   TVSPLLHTPMYYFLSSLSCVDLCYSTVITPKMLVNFLGKKNLIVYSECMA
gi|1083741|    IVSPLLHTPMYYFLSSLSFVDLCYSTVITPKMLVNFLGKKNFITYSECMT
gi|11692559|   NISSQLHSPMYYFLSHLSFIDLCYSSVITPKMLVNFVCAKNTISFKECMT
gi|11692557|   TVSPLLHTPMYYFLSSLSFVDLSYSTVITPKMLVNFLGKKNFITYSECMA
gi|10644515|   TVSPLLHTPMYYFLSSLSCVDLCYSTVITPKMLVNFLGKKNVIVYSECMA

                    110       120       130       140       150
GPCR2          QLFFFVVFVVAEGYLLTAMAYDRYVAICSPLLYNAIMSSWNCSLLVLAAF
gi|11692555|   QLFFFVIFVVAEGYLLTAMAYDRYVAICRPLLYNVIMSSRICSLLVLVAF
gi|1083741|    QFFFFATFVVTEGYLLTAMAYDRYVAICSPLLYNTVMSHKVCSIMMAVVX
gi|11692559|   QLYFFLILATSEGYLLTVMAYDHYVAICRPLLYNVMMSSKHCLLLVLVAF
gi|11692557|   QFFFFAVFVVTEGYLLTVMAYDRYVAICRPLLYNVMMSSKHCLLLVLVAF
gi|10644515|   QLFFFVIFVVAEGYLLTAMAYDRYVAICRPLLYNVIMSSRICSLLVLVAF

                    160       170       180       190       200
GPCR2          FLGFLSALTHTSAMMKLSFCKSHIINHYFCDVLPLLNLSCSNTHLNELLL
gi|11692555|   ILGFVSALAHTSAMMNLSFCKSHVISHYFCDVLPLLNLSCSDIKLNELLL
gi|1083741|    SLGLFSAVVHTSAMMNLSFCKSYIISHYFCDALPLLKLACSNTHLNELLI
gi|11692559|   SLGFFGATVHTTRMTMLSFCGSHIIRHYFCDILPLLTLSCSSTHINEVLL
gi|11692557|   TLGLFSAVVHTSAMMSLNFCKTYIISHYFCDALPLLKLSCSNTHLNELLI
gi|10644515|   ILGFVSALAHTSAMMNLSFCKSHIISHYFCDVLPLLNLSCSNTHLNELLL

                    210       220       230       240       250
GPCR2          FIIAGFNTLVPTLAVAVSYAFILYSILHIRSSEGRSKAFGTCSSHLMAVV
gi|11692555|   FIIAGFNTLVPTLAVAISYVFIFCSILHIKSSKGRSKAFGTCSSHLMAVG
gi|1083741|    FIIGGLNTLVPTLAVAISYVFIFCSILRIRSSEGRSKAFGTCSSHLMAVG
gi|11692559|   FIIGGVNTLAPTLAVTISYAFILTSILRIRSNEGRSKAFGTCSSHIMAVG
gi|11692557|   FIIGGINTLVPTLAVAISYVFIFCSIRHIKSSKSRSKAFGTCSSHLMAVG
gi|10644515|   FIIGGFNTLVPTLAVAISYVFIFCSILHIKSSKGRSKSFGTCSSHLMAVG

                    260       270       280       290       300
GPCR2          IFFGSITFMYFKPPSSNSLDQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|11692555|   IFFGSITFMYFKPPSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|1083741|    IFFGSITFMYLKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|11692559|   IFFGSITFMYFKPPSSNMMEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKT
gi|11692557|   IFFGSITFMYLKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKK
gi|10644515|   IFFGSITFMYFKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSL-------

                    310
GPCR2          ALRKVLVGK---
gi|11692555|   ALGKCLAGR---
```

```
gi|1083741|    ALGRFSVRS---
gi|11692559|   ALKKMVGRRQLS
gi|11692557|   ALGRFSVRR---
gi|10644515|   ------------
```

DOMAIN results for GPCR2 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 2H with the statistics and domain description. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to have significant homology to GPCR2. An alignment of GPCR2 residues 41-290 (SEQ ID NO:6) with 7tm_1 residues 1-254 (SEQ ID NO:34) are shown in Table 2H.

Table 2H. DOMAIN results for GPCR2

PSSMs producing significant alignments: Score   E
                                                                      (bits)   value

gnl|Pfam|pfam00001   7tm_1, 7 transmembrane receptor (rhodopsin family)   85.5   4e-18

```
                      10        20        30        40        50
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          GNLGMILLIAVSPLLHTPMYYFLSSLSFVDFCYSSVITPKMLVNFLGKKN
Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDW

                      60        70        80        90       100
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          TILYSECMVQLFFFVVFVVAEGYLLTAMAYDRYVAICSPLLYNAIMSSWV
Pfam|pfam00001 VFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPRR

                     110       120       130       140       150
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          CSLLVLAAFFLGFLSALTHTSAMMKLSFCKSHIINH----YFCDVLPLLN
Pfam|pfam00001 AKVLILLVWVLALLLSLPP---LLFSWLRTVEEGNT----TVCLIDF---

                     160       170       180       190       200
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          LSCSNTH--LNELLLFIHAGFNTLVPTLAVAVSYAFIL---------Y--
Pfam|pfam00001 -PEESVK--RSYVLLSTLVGF--VLPLLVILVCYTRILRTLRKRARSQ--

                     210       220       230       240       250
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          --------------------------------------------------S
Pfam|pfam00001 --------------------------------------------------R

                     260       270       280       290       300
               ....|....|....|....|....|....|....|....|....|....|
GPCR2          IIHIRSSEGRSKAFGTCSSHLMAVVIFFGSITFMYF----KPPSS-----
Pfam|pfam00001 SLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLDSLCLLSIW-----

                     310       320
               ....|....|....|....|....|.
GPCR2          -NSLDQEKVSSVFYTTVIPMLNPLIY
Pfam|pfam00001 -RVLPTALLIITLWLAYVNSCLNPIIY
```

The nucleic acids and proteins of GPCR2 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further above.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments
5   thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the
10   "Anti-GPCRX Antibodies" section below. The disclosed GPCR2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR2 epitope is from about amino acids 5 to 20. In other specific embodiments, GPCR2 epitopes are from about amino acids 230 to 245, from about amino acids 260 to 275 and from about amino acids 285 to 308

15   **GPCR3**

The disclosed novel GPCR3 nucleic acid (SEQ ID NO:9) of 936 nucleotides (also referred to AP001804_C) is shown in Table 3A. The following genomic clone was identified as having regions with high homology to the GPCR3 homolog. Genomic clone
>acc:AP001804 HTG Homo sapiens chromosome 11 clone RP11-164A10 map 11q,
20   WORKING DRAFT SEQUENCE, in unordered pieces - Homo sapiens, 165058 bp (DNA) was analyzed by Genscan and Grail software to identify exons and putative coding sequences. An ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 934-936. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 3A, and the start and stop
25   codons are in bold letters.

Table 3A. GPCR3 Nucleotide Sequence (SEQ ID NO:9)
ATGGCTGCTGAGAATTCCTCCTTCGTGACACAGTTTATCCTCGCAGGCTTAACTGACCAACCGGGAG
TCCAGATCCCCCTCTTCTTCCTGTTTCTAGGCTTCTACGTGGTCACTGTGGTGGGGAACCTGGGCTT
GATAACCCTGATAAGGCTCAACTCTCACTTGCACACCCCTATGTACTTCTTCCTCTATAACTTGTCC
TTCATAGATTTCTGCTATTCCAGTGTTATCACTCCCAAAATGCTGATGAGCTTTGTCTTAAAGAAGA
ACAGCATCTCCTACGCAGGGTGTATGACTCAGCTCTTCTTCTTTCTTTTTCTTTGTTGTCTCTGAGTC
CTTCATCCTGTCAGCAATGGCGTATGACCGCTATGTGGCCATCTGTAACCCACTGTTGTACATGGTC
ACCATGTCTCCCCAGGTGTGTGTTTTCTCCTTTTGTTGGGTGTCTATGGGATGGGGTTTGCTGGGGCCA
TGGCCCACACAGCCGTGCATGATGGGTGTGACCTTCTGTGCCAATAACCTTGTCAACCACTACATGTG
TGACATCCTTCCCCTTCTTGAGTGTGCTTGCACCAGCACCTATGTGAATGAGCTTGTAGTGTTTGTT
GTTGTGGGCATTGATATTGGTGTGCCCACAGTCACCATCTTCATTTCCTATGCTCTCATTCTCTCCA

```
GCATCTTCCACATTGATTCCACGGAGGGCAGGTCCAAAGCCTTCAGCACCTGCAGCTCCCACATAAT
TGCAGTTTCTCTGTTCTTTGGGTCAGGAGCATTCATGTACCTCAAACCCTTTTCTCTTTTAGCTATG
AACCAGGGCAAGGTGTCTTCCCTATTCTATACCACTGTGGTGCCCATGCTCAACCCATTAATTTATA
GCCTGAGGAATAAGGACGTCAAAGTTGCTCTAAAGAAAATCTTGAACAAAAATGCATTCTCCTGA
```

The GPCR3 protein (SEQ ID NO:10) encoded by SEQ ID NO:9 is 311 aa in length, has a molecular weight of 34480.27 Daltons, and is presented using the one-letter amino acid code in Table 3B. The Psort profile for GPCR3 predicts that these sequences have a signal peptide and are likely to be localized at the plasma membrane with a certainty of 0.600. In alternative embodiments, a GPCR3 polypeptide is located to the Golgi body with a certainty of 0.400, the endoplasmic reticulum (membrane) with a certainty of 0.300, or a microbody (peroxisome) with a certainty of 0.300. The Signal P predicts a likely cleavage site for a GPCR3 peptide is between positions 41 and 42, *i.e.*, at the dash in the sequence VVG-NL.

Table 3B. Encoded GPCR3 protein sequence (SEQ ID NO:10)
MAAENSSFVTQFILAGLTDQPGVQIPLFFLFLGFYVVTVVGNLGLITLIRLNSHLHTPMYFFLYN LSFIDFCYSSVITPKMLMSFVLKKNSISYAGCMTQLFFFLFFVVSESFILSAMAYDRYVAICNPL LYMVTMSPQVCFLLLLGVYGMGFAGAMAHTACMMGVTFCANNLVNHYMCDILPLLECACTSTYVN ELVVFVVVGIDIGVPTVTIFISYALILSSIFHIDSTEGRSKAFSTCSSHIIAVSLFFGSGAFMYL KPFSLLAMNQGKVSSLFYTTVVPMLNPLIYSLRNKDVKVALKKILNKNAFS

The amino acid sequence of GPCR3 had high homology to other proteins as shown in Table 3C.

Table 3C. BLASTX results for GPCRC			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY90877 Human G protein-coupled receptor GTAR11-3	+1	1172	3.2e-118
>patp:AAY90876 Human G protein-coupled receptor GTAR11-2	+1	1143	3.8e-115

In a search of sequence databases, it was found, for example, that the GPCR3 nucleic acid sequence has 659 of 929 bases (70%) identical to a *Rattus norvegicus* Olfactory Receptor-like protein mRNA (GENBANK-ID:AF010293|acc:AF010293). The full GPCR3 amino acid sequence was found to have 234 of 299 amino acid residues (78%) identical to, and 264 of 299 residues (88%) similar to, the 309 amino acid residue Olfactory Receptor-like protein from *Rattus norvegicus* (ptnr:PIR-ID:S29709). GPCR3 has 100% homology to OLFACTORY RECEPTOR 8B8 (OLFACTORY RECEPTOR TPCR85) (gi|14423794|sp|Q15620| O8B8_HUMAN [14423794]) disclosed Apr 21, 2001 on the

GenBank website. See, Vanderhaeghen,*et al.*, 1997 *Genomics* 39 (3), 239-246. GPCR3 also
has homology to the proteins shown in the BLASTP data in Table 3D.

Table 3D. BLAST results for GPCR3					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|14423794|sp|Q156 20|	O8B8_HUMAN OLFACTORY RECEPTOR 8B8 (OLFACTORY RECEPTOR TPCR85)	311	286/311 (91%)	286/311 (91%)	e-134
gi|11692535|gb|AAG3 9864.1|AF282279_1 (AF282279)	odorant receptor K21 [Mus musculus]	310	244/311 (78%)	265/311 (84%)	e-112
gi|11692541|gb|AAG3 9867.1|AF282282_1 (AF282282)	odorant receptor K23 [Mus musculus]	310	231/311 (74%)	258/311 (82%)	e-106
gi|423701|pir||S297 09	olfactory receptor OR14 - rat	304	214/299 (71%)	242/299 (80%)	1e-97
gi|11692539|gb|AAG3 9866.1|AF282281_1 (AF282281)	odorant receptor K22 [Mus musculus]	309	220/311 (70%)	249/311 (79%)	6e-97

5   A multiple sequence alignment is given in Table 3E, with the GPCR3 protein being
shown on line 1 in Table 3E in a ClustalW analysis, and comparing the GPCR3 protein with
the related protein sequences shown in Table 3D. This BLASTP data is displayed graphically
in the ClustalW in Table 3E.

Table 3E. ClustalW Analysis of GPCR3

10   1) GPCR3; SEQ ID NO:10
     2) >gi\|14423794\|sp\|Q15620\|O8B8_ Human Olfactory Receptor 8B8 (OR TPCR85); SEQ ID NO:40
     3) >gi\|11692535\|gb\|AAG39864.1\|AF282279_1 odorant receptor K21 [Mus musculus]; SEQ ID NO:41
     4) >gi\|11692541\|gb\|AAG39867.1\|AF282282_1 odorant receptor K23 [Mus musculus]; SEQ ID NO:42
   5) >gi\|423701\|pir\|\|S29709 olfactory receptor OR14 - rat; SEQ ID NO:43
15   6) >gi\|11692539\|gb\|AAG39866.1\|AF282281_1 odorant receptor K22 [Mus musculus]; SEQ ID NO:44

```
                      10        20        30        40        50
               ....|....|....|....|....|....|....|....|....|....|
   GPCR3       MAAENSSFVTQFILAGLTDQPGVQIPLFFLFLGFYVVTVVGNLGLITLIR
20 gi|14423794| MAAENSSFVTQFILAGLTDQPGVQIPLFFLFLGFYVVTVVGNLGLITLIR
   gi|11692535| MATENA-SVPEFILAGLTDQPGLRMPLFFLFLGFYMVTMVGNLGLITLIG
   gi|11692541| MTAKNS-SVTEFILAGLTDQPGLRMPLFFLFLGFYMVTVVGNLGLISLIG
   gi|423701|   -------SVTEFILAGLTDQPGLRMPLFFLFLGFYMVTVVGNLIGFFLIG
   gi|11692539| MTAKNS-SVTEFILAGLTDQPGLRMPLFFLFLGFYMVTVVGNLGLISLIG

                      60        70        80        90        100
               ....|....|....|....|....|....|....|....|....|....|
   GPCR3       LNSHLHTPMYFFLYNLSFIDFCYSSVITPKMLMSFVLKKNSISYAGCMTQ
30 gi|14423794| LNSHLHTPMYFFLYNLSFIDFCYSSVITPKMLMSFVLKKNSISYAGCMTQ
   gi|11692535| LNSHLHTPMYFFLFNLSLIDFCYSTVITPKMLVSFVSKKNIISYSGCMTQ
   gi|11692541| LNSHLHTPMYFFLFNLSLIDFCYSSTISPKMLMSFISKKNIISHPGCMAQ
   gi|423701|   LNSHLHTPMYFFLFNLSVVDFCFSSTIITPKMLMSFISKKNIISHSGCMTQ
   gi|11692539| LNSHLHTPMYFFLFNLSVIDFCYSSTIIPKMLMNFISRKNIISHSGCMTQ
```

```
                    110       120       130       140       150
           ....|....|....|....|....|....|....|....|....|....|
GPCR3      LFFFLFFVVSESFILSAMAYDRYVAICNPLLYMVTMSPQVCFLLLLGVYG
gi|14423794| LFFFLFFVVSESFILSAMAYDRYVAICNPLLYMVTMSPQVCFLLLLGVYG
gi|11692535| LFFFLFFVVSESFILSAMAYDRYVAICNPLMYTVTMSPQVCLLLLLGVYV
gi|11692541| LFFFCFFVISESFILSAMAYDRYVAICNPLMYMVTMSPQVCLLLFGVYL
gi|423701|   LFFFCFFVVSETFILSAMAYDRYVAICNPLMYTVTMSPQVCLLLLLGAYV
gi|11692539| LFFFCFFVVSESFILSAMAYDRYVAICNPLMYTVTMSPQVYLLLLLGVYV


                    160       170       180       190       200
           ....|....|....|....|....|....|....|....|....|....|
GPCR3      MGFAGAMAHTACMMGVTFCANNLVNHYMCDILPLLECACTSTYVNELVVF
gi|14423794| MGFAGAMAHTACMMGVTFCANNLVNHYMCDILPLLECACTSTYVNELVVF
gi|11692535| MGFAGAMAHTAFMVKLTFCADKLVNHYMCDILPLLERSCTSTYVNELVVF
gi|11692541| MGFVGAMAHTISMARLTFCADNIVNHYMCDILPLLEHSCTSTYVNELVVF
gi|423701|   MGFSEAMAHTGNLMNLTFCADNLVNHFMCDILPLLELSCNSTFINELVVF
gi|11692539| MGFSGAMAIQ-ELMNLTFCADNLINHFMCDILPLLELSCNSTFINELVIF


                    210       220       230       240       250
           ....|....|....|....|....|....|....|....|....|....|
GPCR3      VVVGIDIGVPTVTIFISYALILSSIFHIDSTEGRSKAFSTCSSHIIAVSL
gi|14423794| VVVGIDIGVPTVTIFISYALILSSIFHIDSTEGRSKAFSTCSSHIIAVSL
gi|11692535| IVVGIDIGVPTVTIFISYALILSSILRTSSTEGRSKAFSTCSSHIIAVSL
gi|11692541| IFVSFDIGVPIVTIFISYALILSSILHMHSTEGRSKAFSTCSSHMIVVCL
gi|423701|   IVVAIDIAVPIVSIFISYALILSSILRMHSTEGRSKAFSTCSSHLIVVCL
gi|11692539| IVVAFDIGVPIVTIFISYALILSSILRMHSTEGRSKAFSTYSSHLIVVCL


                    260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|
GPCR3      FFGSGAFMYLKPFSLLAMNQGKVSSLFYTTVVPMLNPLIYSLRNKDVKVA
gi|14423794| FFGSGAFMYLKPFSLLAMNQGKVSSLFYTTVVPMLNPLIYSLRNKDVKVA
gi|11692535| FFGSGAFMYLKPSSLLPMNQGKVSSLFYTIVVPMLNPLIYSLRNKDVKVA
gi|11692541| FFGSGAFMYLQPPSVLSLDQGKVSSLFYTIVVPMLNPLIYSLRNKDVKVA
gi|423701|   LFGSGAFMYLKLPSILPLDQGKVSSLFYTIVVPMLNPLIYSLRNKDVKVA
gi|11692539| FFGSGAFMYLKPPSILPLDQGKVSSLFYMIVVPMLNPLIYSLRNKDVKVG


                    310
           ....|....|.
GPCR3      LKKILNKNAFS
gi|14423794| LKKILNKNAFS
gi|11692535| LRKTLSRSSFS
gi|11692541| VRKTLDRRIFS
gi|423701|   LRKTLGKIILS
gi|11692539| LRKTLGKRILS
```

Table 3F lists the domain description from DOMAIN analysis results against GPCR3. This indicates that the GPCR3 sequence has properties similar to those of other proteins known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID NO:34) itself.

Table 3F Domain Analysis of GPCR3

PSSMs producing significant alignments:

	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	89.0	3e-19

```
                        10        20        30        40        50
                ....|....|....|....|....|....|....|....|....|....|
GPCR3           GNIGLITLIRLNSHIH-TPMYFFLYNLSFIDFCYSSVITPKMLMSFVLK-
Pfam|pfam00001  GNLLVILVILRTKKLR-TPTNIFILNLAVADLLFLLTLPPWALYYLVGG-

                        60        70        80        90        100
                ....|....|....|....|....|....|....|....|....|....|
GPCR3           KNSISYAGCMTQLFFLFFVVSESFILSAMAYDRYVAICNPLLYMVTMSP
Pfam|pfam00001  DWVFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTP

                        110       120       130       140       150
                ....|....|....|....|....|....|....|....|....|....|
GPCR3           QVCFLLLLGVYGMGFAGAMAHTACMMGVTFCANNLVNH----YMCDILPL
Pfam|pfam00001  RRAKVIILLVWVLALLLSLPP---LHFSWLRTVEEGNT----TVCLIDF-

                        160       170       180       190       200
                ....|....|....|....|....|....|....|....|....|....|
GPCR3           LECACTSTYWN---ELVMFVVVGIDIGVPTVTIFISMALIL---------
Pfam|pfam00001  -----PEESVK---RSYVLLSTLVGFVLPLLVILVCYTRILRTLRKRARS

                        210       220       230       240       250
                ....|....|....|....|....|....|....|....|....|....|
GPCR3           S------------SIFHIDSIEGRSKAFSTCSSHLIAVSLFFGSGAFMYL
Pfam|pfam00001  Q------------RSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLL

                        260       270       280
                ....|....|....|....|....|....|....|
GPCR3           KPFSLLA-----MNQGKVSSLFYTTVVPMLNPLIY
Pfam|pfam00001  DSLCLLSIW-RVLPTALLITLWLAYVNSCLNPIIY
```

The nucleic acids and proteins of GPCR3 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further above.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR3 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR3 epitope is from about amino acids 10 to 20. In another embodiment, a GPCR3 epitope is from about amino acids 175 to 190. In specific embodiments, GPCR3 epitopes are from about amino acids 230 to 245and from about amino acids 285 to 311.

27

GPCR4

The disclosed novel GPCR4 nucleic acid (SEQ ID NO:11) of 942 nucleotides (also referred to as AP001804_D) is shown in Table 4A. The following genomic clone was identified as having regions with high homology to GPCR4: genomic clone >acc:AP001804

5   HTG *Homo sapiens* chromosome 11 clone RP11-164A10 map 11q, WORKING DRAFT SEQUENCE, in unordered pieces. The *Homo sapiens* 165058 bp DNA was analyzed by Genscan and Grail software programs to identify exons and putative coding sequences. A GPCR4 ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TAA codon at nucleotides 940-942. The start and stop codons in Table 4A are in bold letters.

10

Table 4A. GPCR4 Nucleotide Sequence (SEQ ID NO:11)
ATGCTGGCTAGAAACAACTCCTTAGTGACTGAATTTATTCTTGCTGGATTAACAGATCATC CAGAGTTCCAGCAACCCCTCTTTTTCCTGTTTCTAGTGGTCTACATTGTCACCATGGTAGG CAACCTTGGCTTGATCATTCTTTTCGGTCTAAATTCTCACCTCCACACACCAATGTACTAT TTCCTCTTCAATCTCTCCTTCATTGATCTCTGTTACTCCTCTGTTTTCACTCCCAAAATGC TAATGAACTTTGTATCAAAAAAGAATATTATCTCCTATGTTGGGTGCATGACTCAGCTGTT TTTCTTTCTCTTTTTTGTCATCTCTGAATGTTACATGTTGACCTCAATGGCATATGATCGC TATGTGGCCATCTGTAATCCATTGCTGTATAAGGTCACCATGTCCCATCAGGTCTGTTCTA TGCTCACTTTTGCTGCTTACATAATGGGATTGGCTGGAGCCACGGCCCACACCGGGTGCAT GCTTAGACTCACCTTCTGCAGTGCTAATATCATCAACCATTACTTGTGTGACATACTCCCC CTCCTCCAGCTTTCCTGCACCAGCACCTATGTCAACGAGGTGGTTGTTCTCATTGTTGTGG GTATTAATATCATGGTACCCAGTTGTACCATCCTCATTTCTTATGTTTTCATTGTCACTAG CATTCTTCATATCAAATCCACTCAAGGAAGATCAAAAGCCTTCAGTACTTGTAGCTCTCAT GTCATTGCTCTGTCTCTGTTTTTTGGGTCAGCGGCATTCATGTATATTAAATATTCTTCTG GATCTATGGAGCAGGGAAAAGTTTCTTCTGTTTTCTACACTAATGTGGTGCCCATGCTCAA TCCTCTCATCTACAGTTTGAGGAACAAGGATGTCAAAGTTGCACTGAGGAAAGCTCTGATT AAAATTCAGAGAAGAAATATATTC**TAA**

The GPCR4 protein (SEQ ID NO:12) encoded by SEQ ID NO:11 has 313 amino acid residues and is presented using the one-letter code in Table 4B. The predicted molecular weight of GPCR4 protein is approximately 35303.38 Daltons. The Psort profile for GPCR4

15   predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6. In alternative embodiments, GPCR4 is located in the Golgi body with a certainty of 0.4, the endoplasmic reticulum (membrane) with a certainty of 0.3 or microbodies (peroxisomes) with a certainty of 0.3. The Signal P predicts a likely cleavage site between positions 44 and 45, i.e., at the dash in the sequence NLG-LI.

Table 4B. Encoded GPCR4 protein sequence (SEQ ID NO:12)
MLARNNSLVTEFILAGLTDHPEFQQPLFFLFLVVYIVTMVGNLGLIILFGLNSHLHTPMYYFLF NLSFIDLCYSSVFTPKMLMNFVSKKNIISYVGCMTQLFFFLFFVISECYMLTSMAYDRYVAICN

```
PLLYKVTMSHQVCSMLTFAAYIMGLAGATAHTGCMLRLTFCSANIINHYLCDILPLLQLSCTST
YVNEVVVLIVVGINIMVPSCTILISYVFIVTSILHIKSTQGRSKAFSTCSSHVIALSLFFGSAA
FMYIKYSSGSMEQGKVSSVFYTNVVPMLNPLIYSLRNKDVKVALRKALIKIQRRNIF
```

The amino acid sequence of GPCR4 had high homology to other proteins as shown in Table 4C.

Table 4C. BLASTX results for GPCR4			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY90877 Human G protein-coupled receptor GTAR11-3	+1	1586	4.3e-162
>patp:AAY90876 Human G protein-coupled receptor GTAR11-2	+1	1544	1.2e-157

5

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention AP001804_D has 650 of 917 bases (70%) identical to a *Rattus norvegicus* Olfactory Receptor-like protein mRNA (GENBANK-

10   ID:AF010293|acc:AF010293). The full amino acid sequence of the protein of the invention was found to have 205 of 300 amino acid residues (68%) identical to, and 247 of 300 residues (82%) similar to, the 309 amino acid residue Olfactory Receptor-like protein from *Rattus norvegicus* (ptnr:PIR-ID:S29709). GPCR4 also has homology to the proteins shown in the BLASTP data in Table 4D.

15

Table 4D. BLAST results for GPCR4					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11692545|gb|AAG3 9869.1|AF282284_1 (AF282284)	odorant receptor K26 [Mus musculus]	314	188/306 (61%)	220/306 (71%)	2e-87
gi|14423794|sp|Q156 20|	O8B8_HUMAN OLFACTORY RECEPTOR 8B8 (OLFACTORY RECEPTOR TPCR85)	311	198/307 (64%)	230/307 (74%)	3e-86
gi|11692537|gb|AAG3 9865.1|AF282280_1 (AF282280)	odorant receptor K21h1 [Mus musculus]	314	189/308 (61%)	227/308 (73%)	5e-86
gi|11692541|gb|AAG3 9867.1|AF282282_1 (AF282282)	odorant receptor K23 [Mus musculus]	310	191/304 (62%)	223/304 (72%)	1e-85
gi|11692535|gb|AAG3 9864.1|AF282279_1 (AF282279)	odorant receptor K21 [Mus musculus]	310	196/304 (64%)	227/304 (74%)	2e-84

A multiple sequence alignment is given in Table 4E, with the GPCR4 protein being shown on line 1 in Table 4E in a ClustalW analysis, and comparing the GPCR4 protein with

the related protein sequences shown in Table 4D. This BLASTP data is displayed graphically

in the ClustalW in Table 4E.

Table 4E. ClustalW Analysis of GPCR4

1) GPCR4; SEQ ID NO:12
2) >gi|11692545|gb|AAG39869.1|AF282284_1 odorant receptor K26 [Mus musculus]; SEQ ID NO: 45
3) >gi|14423794|sp|Q15620|O8B8_ Human Olfactory Receptor 8B8 (OR TPCR85); SEQ ID NO: 46
4) >gi|11692537|gb|AAG39865.1|AF282280_1 odorant receptor K21h1 [Mus musculus]; SEQ ID NO: 47
5) >gi|11692541|gb|AAG39867.1|AF282282_1 odorant receptor K23 [Mus musculus]; SEQ ID NO: 48
6) >gi|11692535|gb|AAG39864.1|AF282279_1 odorant receptor K21 [Mus musculus]; SEQ ID NO: 49

```
                   10        20        30        40        50
             ....|....|....|....|....|....|....|....|....|....|
GPCR4        -----MLARNMSLVTEFILAGLTDHPEFQQPLFFLFLVVYIVTMVGNLGL
gi|11692545| MMLGRMAFSNDSSVKEFILLGLTQQPELQMPLFFLFLGIYVVSMVGNLGL
gi|14423794| -----MAAENSSFVTQFILAGLTDQPGVQIPLFFLFLGFYVVTVVGNLGL
gi|11692537| MSQKRMAPRNSSSVTEFILVGFSNQPALQLPLFFVFLGIYVLTVIGNLGL
gi|11692541| -----MTAKNS-SVTEFILAGLTDQPGLRMPLFFLFLGFYMVTVVGNLGL
gi|11692535| -----MATENA-SVPEFILAGLTDQPGLRMPLFFLFLGFYMVTMVGNLGL

                   60        70        80        90       100
             ....|....|....|....|....|....|....|....|....|....|
GPCR4        IILFGLNSHLHTPMYVFLFNLSFIDLCYSSVRTPKMLMNFVSKKNIISYV
gi|11692545| IVLIVLNPHLHTPMYVFLFNLSFIDLCYSSVITPRMLVGFV-KQNIISHA
gi|14423794| ITLIRLNSHLHTPMYFFLYNLSFIDFCYSSVITPKMLMSFVLKKNSISYA
gi|11692537| ITLIGLNSSLHTPMYFFLFNLSFIDFCYSCVRTPKMLSDFV-SENIISYM
gi|11692541| ISLIGLNSHLHTPMYFFLFNLSLIDFCYSSTISPKMLMSFISKKNIISHP
gi|11692535| ITLIGLNSHLHTPMYFFLFNLSLIDFCYSTVITPKMLVSFVSKKNIISYS

                  110       120       130       140       150
             ....|....|....|....|....|....|....|....|....|....|
GPCR4        GCMTQLFFFLFFVISECYMLTSMAYDRYVAICNPLLYKVTMSHQVCSMLT
gi|11692545| ECMTQLFFFCFFVIDECYILTAMAYDRYAAICKPLLYQVTMSHQVCHLMM
gi|14423794| GCMTQLFFFLFFVVSESFILSAMAYDRYVAICNPLLYMVTMSPQVCFLLL
gi|11692537| GCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYTVTMSPQVCTLLM
gi|11692541| GCMAQLFFFCFFVISESFILSAMAYDRYVAICNPLMYMVTMSPQVCLLLL
gi|11692535| GCMTQLFFFLFFVVSESFILSAMAYDRYVAICNPLMYTVTMSPQVCLLLL

                  160       170       180       190       200
             ....|....|....|....|....|....|....|....|....|....|
GPCR4        FAAYIMGLAGATAHTGCMLRLTFCSANIINHYLCDILPLLQLSCTSTYVN
gi|11692545| VGVYVMGLVGAMAHTGSMLSLTFCDGNIINHYMCDIPPLQKLSCTSTSIN
gi|14423794| LGVYGMGFAGAMAHTACVMGVTFCANNLVNHYMCDILPLLECACTSTYVN
gi|11692537| FCSYVIGFAGAMAHTGSMLTLTFCDSNMIHHYLCEVLPLLQLSCTSTYAN
gi|11692541| FGVYLMGFVGAMAHTISMARLTFCADNIVNHYMCDILPLLEHSCTSTYVN
gi|11692535| LGVYVMGFAGAMAHTAFMVKLTFCADKLVNHYMCDILPLLERSCTSTYVN

                  210       220       230       240       250
             ....|....|....|....|....|....|....|....|....|....|
GPCR4        EVVVLIVVGINIMVPSCTILISYVFIVTSILHIKSTQGRSKAFSTCSSHV
gi|11692545| ELVVFIVVGVNVIIPSLTVFISYTLILSNILSIQSAEGRSKAFSTCGSHV
gi|14423794| ELVVFVVVGIDIGVPTVTIFISYALILSSIFHIDSTEGRSKAFSTCSSHI
gi|11692537| ELVFFIVVGVVITASSISIFISYALILSNILKIPSAEGRSKAFGTCGSHV
gi|11692541| ELVVFIFVSFDIGVPIVTIFISYALILSSILHMHSTEGRSKAFSTCSSHM
gi|11692535| ELVVFIVVGIDIGVPTVTIFISYALILSSILRISSTEGRSKAFSTCSSHI

                  260       270       280       290       300
```

```
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         IALSLFFGSAAFMYIKYS-SGSMEQGKVSSVFYTNVVPMLNPLIYSLRNK
gi|11692545|  IAVSLFFGASAFMYLKPS-SASVDDDKISTIFYTIVGPMLNPFIYSLRNK
gi|14423794|  IAVSLFFGSGAFMYLKPFSLLAMNQGKVSSLFYTTVVPMLNPLIYSLRNK
5   gi|11692537|  VAVALFFGSGAFTYLTTSFPGSMEEGRFASVFYTNVVPMLNPLIYSLRNK
gi|11692541|  IVVCLFFGSGAFMYLQPPSVLSLDQGKVSSLFYTIVVPMLNPLIYSLRNK
gi|11692535|  IAVSLFFGSGAFMYLKPSSLLPMNQGKVSSLFYTIVVPMLNPLIYSLRNK


                            310       320
              ....|....|....|....|....|
10  GPCR4         DVKVALRKALIKIQRRNIF-
gi|11692545|  DVHIALRKTLKKS----MFI
gi|14423794|  DVKVALKKILNKN----AFS
gi|11692537|  DVKLALNKTLKRV----LF-
15  gi|11692541|  DVKVAVRKTLDRR----LFS
gi|11692535|  DVKVALRKTLSRS----SFS
```

Table 4F lists the domain description from DOMAIN analysis results against GPCR4. This indicates that the GPCR4 sequence has properties similar to those of other proteins known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID NO:34) itself.

Table 4F Domain Analysis of GPCR4

PSSMs producing significant alignments:	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	82.8	2e-17

```
                         10        20        30        40        50
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         GNLGLILFGLNSHIHTPMYYFLFNLSFIDLCYSSVFTPKMLMNFVSK-K
Pfam|pfam00001  GNLLVILVILRTKKLRTPTNIFILNLAVADLLFLLTLPPWALYYLVGG-D


                         60        70        80        90       100
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         NHISYVGCMTQLFFLFFVISECYMLTSMAYDRYVAICNPLLYKVTMSHQ
Pfam|pfam00001  WVFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPR


                        110       120       130       140       150
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         VCSMLTFAAYIMGLAGATAH-TGCMLRLTFCSANIINHYLCDILPLLQLS
Pfam|pfam00001  RAKVLILLVWVIALLLSLPPLLFSWLRTVEEG----NTTVCLIDFPEESV


                        160       170       180       190       200
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         C-----TSTYVNEVVVLIVVGINIMVPSCTILISYVFIV---------T-
Pfam|pfam00001  K-----RSYV------LLSTLVGFVLPLLVIIVCYTRILRTLRKRARSQ-


                        210       220       230       240       250
              ....|....|....|....|....|....|....|....|....|....|
GPCR4         -----------SILHIKSTQGRSKAFSTCSSHVIALSLFFGSAAFMYIKY
Pfam|pfam00001  ----------RSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLDS


                        260       270       280
              ....|....|....|....|....|....|....|
GPCR4         -----SS-GSMEQGKVSSVFYTNVVPMLNPLIY
```

Pfam|pfam00001   LCLLSIW-RV{PTAL}I{TLW}LAY{V}NSC{LNPIIY}

      The GPCR4 protein predicted here is similar to the "Olfactory Receptor-Like Protein Family", some members of which end up localized at the cell surface where they exhibit

5   activity. Therefore, it is likely that this novel GPCR4 protein is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application. The Olfactory Receptor-like GPCR4 proteins disclosed are expressed in at least the following tissues: olfactory receptors are a class of G protein-coupled receptor which are known to be expressed in all tissue types especially olfactory epithelium. Further tissue

10   expression analysis is provided in the Examples.

      The nucleic acids and proteins of GPCR4 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further herein.

      These materials are further useful in the generation of antibodies that bind

15   immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR4 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR4

20   epitope is from about amino acids 5 to 20. In additional embodiments, GPCR4 epitopes are from about amino acids 225 to 245, from about amino acids 260 to 275 and from about amino acids 290 to 313.

GPCR5

      A second GPCR-like protein of the invention, referred to herein as GPCR5, is an

25   Olfactory Receptor ("OR")-like protein. The novel GPCR5 nucleic acid sequences were identified on chromosome 11 as described in Example 1. Some members of the Olfactory Receptor-Like Protein Family end up localized at the cell surface, where they exhibit activity. Therefore it is likely that these novel GPCR5 proteins are available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

30   Two alternative novel GPCR5 nucleic acids and encoded polypeptides are provided, namely GPCR5a and GPCR5b.

GPCR5a

In one embodiment, a GPCR5 variant is the novel GPCR5a (alternatively referred to herein as CG56040_01), which includes the 912 nucleotide sequence (SEQ ID NO:13) shown in Table 5A. The DNA sequence and protein sequence for GPCR5a or one of its splice forms

5   was obtained solely by exon linking. A GPCR5a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 910-9120, shown in bold in Table 5A.

Table 5A. GPCR5 Nucleotide Sequence (SEQ ID NO:13)
ATGACTCTGAGAAACAGCTCCTCAGTGACTGAGTTTATCCTTGTGGGATTATCAGAACAGC CAGAGCTCCAGCTCCCTCTTTTCCTTCTATTCTTAGGGATCTATGTGTTCACTGTGGTGGG CAACTTGGGCTTGATCACCTTAATTGGGATAAATCCTAGCCTTCACACCCCCATGTACTTT TTCCTCTTCAACTTGTCCTTTATAGATCTCTGTTATTCCTGTGTGTTTACCCCCAAAATGC TGAATGACTTTGTTTCAGAAAGTATCATCTCTTATGTGGGATGTATGACTCAGCTATTTTT CTTCTGTTTCTTTGTCAATTCTGAGTGCTATGTGTTGGTATCAATGGCCTATGATCGCTAT GTGGCCATCTGCAACCCCCTGCTCTACATGGTCACCATGTCCCCAAGGGTCTGCTTTCTGC TGATGTTTGGTTCCTATGTGGTAGGGTTTGCTGGGGCCATGGCCCACACTGGAAGCATGCT GCGACTGACCTTCTGTGATTCCAACGTCATTGACCATTATCTGTGTGACGTTCTCCCCCTC TTGCAGCTCTCCTGCACCAGCACCCATGTCAGTGAGCTGGTATTTTTCATTGTTGTTGGAG TAATCACCATGCTATCCAGCATAAGCATCGTCATCTCTTACGCTTTGATACTCTCCAACAT CCTCTGTATTCCTTCTGCAGAGGGCAGATCCAAAGCCTTTAGCACATGGGGCTCCCACATA ATTGCTGTTGCTCTGTTTTTTGGGTCAGGGACATTCACCTACTTAACAACATCTTTTCCTG GCTCTATGAACCATGGCAGATTTGCCTCAGTCTTTTACACCAATGTGGTTCCCATGCTTAA CCCTTCGATCTACAGTTTGAGGAATAAGGATGATAAACTTGCCCTGGGCAAACCC**TGA**

10 The sequence of GPCR5a was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. The cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was based on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion

15 thereof.

The cDNA coding for the GPCR5a sequence was cloned by the polymerase chain reaction (PCR). Primers were designed based on *in silico* predictions of the full length or some portion (one or more exons) of the cDNA/protein sequence of the invention. The DNA sequence and protein sequence for a novel Olfactory Receptor-like gene were obtained by

20 exon linking and are reported here as GPCR5a. These primers and methods used to amplify GPCR5 a cDNA are described in the Examples.

The GPCR5a polypeptide (SEQ ID NO:14) encoded by SEQ ID NO:13 is 303 aa in length, has a molecular weight of 33640.94 Daltons, and is presented using the one-letter amino acid code in Table 5B. The Psort profile for both GPCR5a and GPCR5b predicts that

these sequences have a signal peptide and are likely to be localized at the plasma membrane with a certainty of 0.600. In alternative embodiments, a GPCR5 polypeptide is located to the mitochondrial inner membrane with a certainty of 0.4195, the Golgi body with a certainty of 0.400, or the mitochondrial intermembrane space with a certainty of 0.3631. The Signal P predicts a likely cleavage site for a GPCR5 peptide is between positions 41 and 42, *i.e.*, at the dash in the sequence VVG-NL.

Table 5B. GPCR5a protein sequence (SEQ ID NO:14)
MTLRNSSSVTEFILVGLSEQPELQLPLFLLFLGIYVFTVVGNLGLITLIGINPSLHTPMYFFLFNLS FIDLCYSCVFTPKMLNDFVSESIISYVGCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYMVT MSPRVCFLLMFGSYVVGFAGAMAHTGSMLRLTFCDSNVIDHYLCDVLPLLQLSCTSTHVSELVFFIV VGVITMLSSISIVISYALILSNILCIPSAEGRSKAFSTWGSHIIAVALFFGSGTFTYLTTSFPGSMN HGRFASVFYTNVVPMLNPSIYSLRNKDDKLALGKP

GPCR5b

In an alternative embodiment, a GPCR5 variant is the novel GPCR5b (alternatively referred to herein as AP001804_B), which includes the 930 nucleotide sequence (SEQ ID NO:15) shown in Table 5C. The GPCR5b ORF begins with a Kozak consensus ATG initiation codon at nucleotides 1-3 and ends with a TAA codon at nucleotides 928-930, which are in bold letters in Table 5C.

Table 5C. GPCR5b Nucleotide Sequence (SEQ ID NO:15)
ATGACTCTGAGAAACAGCTCCTCAGTGACTGAGTTTATCCTTGTGGGATTATCAGAACAGC CAGAGCTCCAGCTCCCTCTTTTCCTTCTATTCTTAGGGATCTATGTGTTCACTGTGGTGGG CAACTTGGGCTTGATCACCTTAATTGGGATAAATCCTAGCCTTCACACCCCCATGTACTTT TTCCTCTTCAACTTGTCCTTTATAGATCTCTGTTATTCCTGTGTGTTTACCCCCAAAATGC TGAATGACTTTGTTTCAGAAAGTATCATCTCTTATGTGGGATGTATGACTCAGCTATTTTT CTTCTGTTTCTTTGTCAATTCTGAGTGCTATGTGTTGGTATCAATGGCCTATGATCGCTAT GTGGCCATCTGCAACCCCCTGCTCTACATGGTCACCATGTCCCCAAGGGTCTGCTTTCTGC TGATGTTTGGTTCCTATGTGGTAGGGTTTGCTGGGGCCATGGCCCACACTGGAAGCATGCT GCGACTGACCTTCTGTGATTCCAACGTCATTGACCATTATCTGTGTGACGTTCTCCCCCTC TTGCAGCTCTCCTGCACCAGCACCCATGTCAGTGAGCTGGTATTTTTCATTGTTGTTGGAG TAATCACCATGCTATCCAGCATAAGCATCGTCATCTCTTACGCTTTGATACTCTCCAACAT CCTCTGTATTCCTTCTGCAGAGGGCAGATCCAAAGCCTTTAGCACATGGGGCTCCCACATA ATTGCTGTTGCTCTGTTTTTTGGGTCAGGGACATTCACCTACTTAACAACATCTTTTCCTG GCTCTATGAACCATGGCAGATTTGCCTCAGTCTTTTACACCAATGTGGTTCCCATGCTTAA CCCTTCGATCTACAGTTTGAGGAATAAGGATGATAAACTTGCCCTGGGCAAAACCCTGAAG AGAGTGCTCTTCTAA

The GPCR5b protein (SEQ ID NO:16) encoded by SEQ ID NO:15 is 309 amino acid in length, has a molecular weight of 34401.88 Daltons, and is presented using the one-letter code in Table 5D. As with GPCR5a, the most likely cleavage site for a GPCR5b peptide is

between amino acids 41 and 42, *i.e.*, at the dash in the sequence VVG-NL, based on the

SignalP result.

Table 5D. GPCR5b protein sequence (SEQ ID NO:16)
MTLRNSSSVTEFILVGLSEQPELQLPLFLLFLGIYVFTVVGNLGLITLIGINPSLHTPMYFFLFNLS FIDLCYSCVFTPKMLNDFVSESIISYVGCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYMVT MSPRVCFLLMFGSYVVGFAGAMAHTGSMLRLTFCDSNVIDHYLCDVLPLLQLSCTSTHVSELVFFIV VGVITMLSSISIVISYALILSNILCIPSAEGRSKAFSTWGSHIIAVALFFGSGTFTYLTTSFPGSMN HGRFASVFYTNVVPMLNPSIYSLRNKDDKLALGKTLKRVLF

5   **GPCR5 Clones**

Unless specifically addressed as GPCR5a or GPCR5b, any reference to GPCR5 is

assumed to encompass all variants. The GPCR5 nucleic acid sequences have alternative 3'

sequences: *i.e.*, GPCR5b has an A residue inserted at position 907 and extends 17 bp beyond

the 3' end of GPCR1a. The GPCR5 polypeptides have alternative carboxyterminal sequences

10   beginning at residue 303, wherein GPCR5 has a proline (P) as a terminal residue at position

303, and GPCR5b contains the sequence TLKRVLF at positions 303-309.

The amino acid sequence of GPCR2 had high homology to other proteins as shown in

Table 5E.

Table 5E. BLASTX results for GPCR5			
Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Prob P(N)
>patp:AAY90878 Human G protein-coupled receptor GTAR11-4	+1	1594	6.1e-163

15

The GPCR disclosed in this invention maps to chromosome 11q25. This information

was assigned using OMIM, the electronic northern bioinformatics tool implemented by

CuraGen Corporation, public ESTs, public literature references and/or genomic clone

20   homologies. This was executed to derive the chromosomal mapping of the SeqCalling

assemblies, Genomic clones, literature references and/or EST sequences that were included in

the invention.

In a search of sequence databases, it was found, for example, that the GPCR5a nucleic

acid sequence has 537 of 706 bases (76%) identical to a gb:GENBANK-

25   ID:AF065872|acc:AF065872.1 mRNA from Homo sapiens (Homo sapiens OR8C1P

pseudogene, partial sequence). The full amino acid sequence of the protein of the invention

was found to have 187 of 296 amino acid residues (63%) identical to, and 233 of 296 amino

acid residues (78%) similar to, the 304 amino acid residue ptnr:SPTREMBL-ACC:Q9QW36

protein from Rattus sp (OR14=ODORANT RECEPTOR In further a search of sequence

databases, it was found, for example, that the GPCR5b nucleic acid sequence has 640 of 926

bases (69%) identical to a *Rattus norvegicus* Olfactory Receptor-like protein mRNA

(GENBANK-ID:AF010293|acc:AF010293). The full amino acid sequence of the protein of

5   the invention was found to have 189 of 302 amino acid residues (62%) identical to, and 238

of 302 residues (78%) similar to, the 309 amino acid residue Olfactory Receptor-like protein

from *Rattus norvegicus* (ptnr:PIR-ID:S29709). Additional BLAST results are shown in Table

5F.

Table 5F. BLAST results for GPCR5					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11692537|gb|AAG3 9865.1|AF282280_1 (AF282280)	odorant receptor K21h1 [Mus musculus]	314	256/309 (82%)	270/309 (86%)	e-113
gi|11692545|gb|AAG3 9869.1|AF282284_1 (AF282284)	odorant receptor K26 [Mus musculus]	314	196/309 (63%)	235/309 (75%)	7e-87
gi|11692543|gb|AAG3 9868.1|AF282283_1 (AF282283)	odorant receptor K25 [Mus musculus]	309	187/309 (60%)	229/309 (73%)	7e-85
gi|11692541|gb|AAG3 9867.1|AF282282_1 (AF282282)	odorant receptor K23 [Mus musculus]	310	195/310 (62%)	235/310 (74%)	2e-84
gi|11692535|gb|AAG3 9864.1|AF282279_1 (AF282279)	odorant receptor K21 [Mus musculus]	310	199/305 (65%)	235/305 (76%)	7e-84

10

A multiple sequence alignment is given in Table 5G, with the GPCR5 protein of the

invention being shown on line 1, in a ClustalW analysis comparing GPCR5 with related

protein sequences, shown in Table 5F. The residue that differs between GPCR5a and GPCR5b

is marked with the (o) symbol.

15   **Table 5G. Information for the ClustalW proteins:**

1. GPCR5b; SEQ ID NO:16
2. gi|11692537|gb|AAG39865.1|AF282280_1 odorant receptor K21h1 [Mus musculus]; SEQ ID NO:50
3. gi|11692545|gb|AAG39869.1|AF282284_1 odorant receptor K26 [Mus musculus]; SEQ ID NO:51
4. gi|11692543|gb|AAG39868.1|AF282283_1 odorant receptor K25 [Mus musculus]; SEQ ID NO:52
20   5. gi|11692541|gb|AAG39867.1|AF282282_1 odorant receptor K23 [Mus musculus]; SEQ ID NO:53
6. gi|11692535|gb|AAG39864.1|AF282279_1 odorant receptor K21 [Mus musculus]; SEQ ID NO:54

```
                      10        20        30        40        50
                 ....|....|....|....|....|....|....|....|....|....|
25   GPCR5       -----MTLRNSSSVTEFILVGLSEQPELQLPLFLLFLGIYVFTVVGNLGL
     gi|11692537|MSQKRMAPRNSSSVTEFILVGFSNQPALQLPLFFVFLGIYVLTVIGNLGL
     gi|11692545|MMLGRMAFSNDSSVKEFILLGLTQQPELQMPLFFLFLGIYVVSMVGNLGL
     gi|11692543|-----MGFGNDSSVKEFILLGLTQQPELQLPLFFLFLGIYVVSIVGNLGL
     gi|11692541|-----MTAKN-SSVTEFILAGLTDQPGLRMPLFFLFLGFYMVTVVGNLGL
30   gi|11692535|-----MATEN-ASVPEFILAGLTDQPGLRMPLFFLFLGFYMVTMVGNLGL
```

```
                        60        70        80        90       100
                   ....|....|....|....|....|....|....|....|....|....|
        GPCR5      ITLIGINPSLHTPMYFFLFNLSFIDLCYSCVFTPKMLNDFVS-ESIISYV
    5   gi|11692537|ITLIGLNSSLHTPMYFFLFNLSFIDFCYSCVFTPKMLSDFVS-ENIISYM
        gi|11692545|IVLIVLNPHLHTPMYYFLFNLSFIDLCYSSVITPRMLVGFVK-QNIISHA
        gi|11692543|IVLIVLNPHLHTPMYYFLFNLSFVDFCYSSVITPKMLVSFVT-QNIISHA
        gi|11692541|ISLIGLNSHLHTPMYFFLFNLSLIDFCYSSTISPKMLMSFISKKNIISHP
        gi|11692535|ITLIGLNSHLHTPMYFFLFNLSLIDFCYSTVITPKMLVSFVSKKNIISYS

                       110       120       130       140       150
                   ....|....|....|....|....|....|....|....|....|....|
   10   GPCR5      GCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYMVTMSPRVCFLLM
        gi|11692537|GCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYIVTMSPQVCTLLM
   15   gi|11692545|ECMTQLFFFCFFVLDECYILTAMAYDRYAAICKPLLYQVTMSHQVCHLMM
        gi|11692543|ECMTQLFFFAFFVIDECYILTAMAYDRYAAICKPLLYQVTMSHQVCHFMM
        gi|11692541|GCMAQLFFFCFFVISESFILSAMAYDRYVAICNPLMYMVTMSPQVCLLLI
        gi|11692535|GCMTQLFFFLFFVVSESFILSAMAYDRYVAICNPLMYIVTMSPQVCLLLI

                       160       170       180       190       200
                   ....|....|....|....|....|....|....|....|....|....|
   20   GPCR5      FCSYVVGFAGAMAHTGSMLRLTFCDSNVIDHYLCDVLPLLQLSCTSTHVS
        gi|11692537|FCSYVIGFAGAMAHTGSMLTLTFCDSNMIHHYLCEVLPLLQLSCTSTYAN
        gi|11692545|VGVYVMGLVGAMAHTGSMLSLTFCDGNIINHYMCDIPPLQKLSCTSTSIN
        gi|11692543|MGVVVMGSVGAVAHIICMLRLTFCDGNIINHYMCDIPPLLKLSCTSTYIN
   25   gi|11692541|FGVYLMGFVGAMAHTISMARLTFCADNIVNHYMCDILPLLEHSCTSTYVN
        gi|11692535|LGVYVMGFAGAMAHTAFMVKLTFCADKLVNHYMCDILPLLERSCTSTYVN

                       210       220       230       240       250
                   ....|....|....|....|....|....|....|....|....|....|
   30   GPCR5      ELVFFIVVGVITMLSSISTVISYALILSNILCIPSAEGRSKAFSTWGSHI
        gi|11692537|ELVFFIVVGVVITASSISIFISYALILSNILKIPSAEGRSKAFGTCGSHV
        gi|11692545|ELVVFIVVGVNVILPSLTVFISYTLILSNILSIQSAEGRSKAFSTCGSHV
        gi|11692543|ELVVFIIVGFNVIVPILTIFISYTLILSNILSIHSAEGRSKAFSTCGSHV
        gi|11692541|ELVVFIFVSFDIGVPIVTIFISYALILSSILHMHSTEGRSKAFSTCSSHM
   35   gi|11692535|ELVVFIVVGIDIGVPFVTIFISYALILSSILRISSTEGRSKAFSTCSSHI

                       260       270       280       290       300
                   ....|....|....|....|....|....|....|....|....|....|
        GPCR5      IAVALFFGSGTFTYLTTSFPGSMNHGRFASVFYTNVVPMLNPSIYSLRNK
   40   gi|11692537|VAVALFFGSGAFTYLTTSFPGSMEEGRFASVFYTNVVPMLNPLIYSLRNK
        gi|11692545|IAVSLFFCASAFMYLKP-SSASVDDDKISTIFYTIVGPMLNPFIYSLRNK
        gi|11692543|IAVSIFFGSLAFMYLKP-SSASVDDDKISTIFYTIVGPMLNPFIYSLRNK
        gi|11692541|IVVCLFFGSGAFMYLQPPSVLSLDQGKVSSLFYTIVVPMLNPLIYSLRNK
        gi|11692535|IAVSLFFGSGAFMYLKPSSLLPMNQGKVSSLFYTIVVPMLNPLIYSLRNK

                       310
                   ....|....|....|.
                       ooooooo
        GPCR5      DDKLALGKTLKRVLE-
   50   gi|11692537|DVKLALNKTLKRVLE-
        gi|11692545|DVHIALRKTLKKSMEI
        gi|11692543|DVHIAMRKTLKKGMFA
        gi|11692541|DVKVAVRKTLDRRIES
        gi|11692535|DVKVALRKTLSRSSFS
   55
```

DOMAIN results for GPCR5 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 5H with the statistics and domain description. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to

have significant homology to GPCR5. An alignment of GPCR5 residues 41-289 with 7tm_1 residues 1-254 (SEQ ID NO:34) are shown in Table 5H.

Table 5H. DOMAIN results for GPCR5

PSSMs producing significant alignments:

	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	93.2	2e-20

```
                 10        20        30        40        50
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      GNLGLITLIGINPSLHTPMYFFLFNLSFIDICYSCVFTPKMLNDFVSES-
Pfam|pfam00001  GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDW

                 60        70        80        90        100
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      IISYVGCMTQLFFFCFFVNSECYVLVSMAYDRYVAICNPLLYMVTMSPRV
Pfam|pfam00001  VFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPRR

                 110       120       130       140       150
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      CFLLMFGSYVVGFAGAMAHTGSMLRLTFCDSNV----IDHYLCDVLPLLQ
Pfam|pfam00001  AKVLILLLVWVLALLLSLPPLLFSWLRTVEEGNT----TVCLIDFPEESVK

                 160       170       180       190       200
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      LSCTSTHVSELVFFIVVGVITMLSSISIVLSYALILS-------------
Pfam|pfam00001  RSYVLLSTLVGFVLPLLVILVCYTRILRTLRKRARSQ-------------

                 210       220       230       240       250
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      -------------------------------------------------
Pfam|pfam00001  -------------------------------------------------

                 260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      ------------------------------------------NILCIPSAEGRS
Pfam|pfam00001  ------------------------------------------RSLKRRSSSERK

                 310       320       330       340       350
           ....|....|....|....|....|....|....|....|....|....|
GPCR5      KAFSTWGSHLIAVALKFGSGTFTYL----TTSFP-GSMNHGRFASVFYTN
Pfam|pfam00001  AAKMLLVVVVVVFVLCWLPYHIVLLLDSLCLLSIW-RVLPTALLITLWLAY

                 360
           ....|....|
GPCR5      VVPMLNPSIY
Pfam|pfam00001  VNSCLNPIIY
```

The nucleic acids and proteins of GPCR5 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further above.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the

generation of antibodies that bind immunospecifically to the novel substances of the invention

for use in therapeutic or diagnostic methods. These antibodies may be generated according to

methods known in the art, using prediction from hydrophobicity charts, as described in the

"Anti-GPCRX Antibodies" section below. The disclosed GPCR5 protein has multiple

5   hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a

contemplated GPCR5 epitope is from about amino acids 5 to 20. In other specific

embodiments, GPCR5 epitopes are from about amino acids 230 to 245, from about amino

acids 255 to 275 and from about amino acids 285 to 309.

GPCR6

10   A further GPCR-like protein of the invention, referred to herein as GPCR6, is an

Olfactory Receptor ("OR")-like protein. The novel GPCR6 nucleic acid sequences were

identified on chromosome 11 as described in Example 1. Some members of the Olfactory

Receptor-Like Protein Family end up localized at the cell surface, where they exhibit activity.

Therefore it is likely that these novel GPCR6 proteins are available at the appropriate sub-

15   cellular localization and hence accessible for the therapeutic uses described in this application.

          Two alternative novel GPCR6 nucleic acids and encoded polypeptides are provided,

namely GPCR6a and GPCR6b.

GPCR6a

          In one embodiment, a GPCR6 variant is the novel GPCR6a (alternatively referred to

20   herein as CG56025-01), which includes the 971 nucleotide sequence (SEQ ID NO:17) shown

in Table 6A. A GPCR6a ORF begins with a Kozak consensus ATG initiation codon at

nucleotides 20-22 and ends with a TAA codon at nucleotides 956-958. Putative untranslated

regions upstream from the initiation codon and downstream from the termination codon are

underlined in Table 6A, and the start and stop codons are in bold letters.

25

Table 6A. GPCR6 Nucleotide Sequence (SEQ ID NO:17)
TTCTAGGGTATCAAGGGACATGAGAAATGGCACAGTAATCACAGAATTCATCCTGCTAGGCTTTCCT
GTTATCCAAGGCCTACAAACACCTCTCTTTATTGCAATCTTTCTCACCTACATATTAACCCTTGCAG
GCAATGGGCTTATTATTGCCACTGTGTGGGCTGAGCCCAGGCTACAAATTCCAATGTACTTCTTCCT
TTGTAACTTGTCTTTCTTAGAAATCTGGTACACCACCACAGTCATCCCCAAACTGCTAGGAACCTTT
GTAGTGGCAAGAACAGTAATCTGCATGTCCTGCTGCCTGCTGCAGGCCTTCTTCCACTTCTTCGTGG
GCACCACCGAGTTCTTGATCCTCACTATCATGTCTTTTGACCGCTACCTCACCATCTGCAATCCCCT
TCACCACCCCACCATCATGACCAGCAAACTCTGCCTGCAGCTGGCCCTGAGCTCCTGGGTGGTGGGC
TTCACCATTGTCTTTTGTCAGACGATGCTGCTCATCCAGTTGCCATTCTGTGGCAATAATGTTATCA
GTCATTTCTACTGTGATGTTGGGCCCAGTTTGAAAGCCGCCTGCATAGACACCAGCATTTTGGAACT
CCTGGGCGTCATAGCAACCATCCTTGTGATCCCAGGGTCACTTCTCTTTAATATGATTTCTTATATC
TACATTCTGTCCGCAATCCTACGAATTCCTTCAGCCACTGGCCACCAAAAGACTTTCTCTACCTGTG
CCTCGCACCTGACAGTTGTCTCCCTGCTCTACGGGGCTGTTCTGTTCATGTACCTAAGACCCACAGC

```
ACACTCCTCCTTTAAGATTAATAAGGTGGTGTCTGTGCTAAATACTATCCTCACCCCCCTTCTGAAT
CCCTTTATTTATACTATTAGAAACAAGGAGGTGAAGGGAGCCTTAAGAAAGGCAATGACTTGCCCAA
AGACTGGTCATGCAAAGTAAAACATGCAACACA
```

     The sequence of GPCR6a was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. The cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was based

5   on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

     The cDNA coding for the GPCR6a sequence was cloned by the polymerase chain reaction (PCR). Primers were designed based on *in silico* predictions of the full length or some

10   portion (one or more exons) of the cDNA/protein sequence of the invention. The DNA sequence and protein sequence for a novel Olfactory Receptor-like gene were obtained by exon linking and are reported here as GPCR6a.These primers and methods used to amplify GPCR6 a cDNA are described in the Examples.

     The GPCR6a polypeptide (SEQ ID NO:18) encoded by SEQ ID NO:17 is 312 aa in

15   length, has a molecular weight of 34526.32 Daltons, and is presented using the one-letter amino acid code in Table 6B. The Psort profile for both GPCR6a and GPCR6b predicts that these sequences have a signal peptide and are likely to be localized at the endoplasmic reticulum (membrane) with a certainty of 0.685. In alternative embodiments, a GPCR6 polypeptide is located to the plasma membrane with a certainty of 0.6400, the Golgi body

20   with a certainty of 0.4600, or to endoplamic reticulum (lumen) with a certainty of 0.1000. The Signal P predicts a likely cleavage site for a GPCR6 peptide is between positions 49 and 50, *i.e.*, at the dash in the sequence VWA-EP.

Table 6B. GPCR6a protein sequence (SEQ ID NO:18)
MRNGTVITEFILLGFPVIQGLQTPLFIAIFLTYILTLAGNGLIIATVWAEPRLQIPMYFFLCNLSFL EIWYTTTVIPKLLGTFVVARTVICMSCCLLQAFFHFFVGTTEFLILTIMSFDRYLTICNPLHHPTIM TSKLCLQLALSSWVVGFTIVFCQTMLLIQLPFCGNNVISHFYCDVGPSLKAACIDTSILELLGVIAT ILVIPGSLLFNMISYIYILSAILRIPSATGHQKTFSTCASHLTVVSLLYGAVLFMYLRPTAHSSFKI NKVVSVLNTILTPLLNPFIYTIRNKEVKGALRKAMTCPKTGHAK

GPCR6b

25   In an alternative embodiment, a GPCR6 variant is the novel GPCR6b (alternatively referred to herein as AP001804_B), which includes the 939 nucleotide sequence (SEQ ID NO:19) shown in Table 6C. The GPCR6b ORF begins with a Kozak consensus ATG

initiation codon at nucleotides 1-3 and ends with a TAA codon at nucleotides 937-939, which are in bold letters in Table 6C. The GPCR6 protein encoded by SEQ ID NO:7 is identical to SEQ ID NO:18, above.

Table 6C. GPCR6b Nucleotide Sequence (SEQ ID NO:19)
ATGAGAAATGGCACAGTAATCACAGAATTCATCCTGCTAGGCTTTCCTGTTATCCAAGGCC
TACAAACACCTCTCTTTATTGCAATCTTTCTCACCTACATATTAACCCTTGCAGGCAATGG
GCTTATTATTGCCACTGTGTGGGCTGAGCCCAGGCTACAAATTCCAATGTACTTCTTCCTT
TGTAACTTGTCTTTCTTAGAAATCTGGTACACCACCACAGTCATCCCCAAACTGCTAGGAA
CCTTTGTAGTGGCAAGAACAGTAATCTGCATGTCCTGCTGCCTGCTGCAGGCCTTCTTCCA
CTTCTTCGTGGGCACCACCGAGTTCTTGATCCTCACTATCATGTCTTTTGACCGCTACCTC
ACCATCTGCAATCCCCTTCACCACCCCACCATCATGACCAGCAAACTCTGCCTGCAGCTGG
CCCTGAGCTCCTGGGTGGTGGGCTTCACCATTGTCTTTTGTCAGACGATGCTGCTCATCCA
GTTGCCATTCTGTGGCAATAATGTTATCAGTCATTTCTACTGTGATGTTGGGCCCAGTTTG
AAAGCCGCCTGCATAGACACCAGCATTTTGGAACTCCTGGGCGTCATAGCAACCATCCTTG
TGATCCCAGGGTCACTTCTCTTTAATATGATTTCTTATATCTACATTCTGTCCGCAATCCT
ACGAATTCCTTCAGCCACTGGCCACCAAAAGACTTTCTCTACCTGTGCCTCGCACCTGACA
GTTGTCTCCCTGCTCTACGGGGCTGTTCTGTTCATGTACCTAAGACCCACAGCACACTCCT
CCTTTAAGATTAATAAGGTGGTGTCTGTGCTAAATACTATCCTCACCCCCCTTCTGAATCC
CTTTATTTATACTATTAGAAACAAGGAGGTGAAGGGAGCCTTAAGAAAGGCAATGACTTGC
CCAAAGACTGGTCATGCAAAGTAA

GPCR6 Clones

Unless specifically addressed as GPCR6a or GPCR6b, any reference to GPCR6 is assumed to encompass all variants. Residue differences between any GPCRX variant sequences herein are written to show the residue in the "a" variant, the residue position with respect to the "a" variant, and the residue in the "b" variant. For example, the GPCR6 nucleic acid sequences differ at the following position: G394T.

In a search of sequence databases, it was found, for example, that the GPCR6 nucleic acid sequence of this invention has 770 of 922 bases (83%) identical to a *Rattus norvegicus* Olfactory Receptor-like protein mRNA (GENBANK-ID:RNOLP4|acc:X80671). The full amino acid sequence of the protein of the invention was found to have 247 of 302 amino acid residues (81%) identical to, and 261 of 302 residues (86%) similar to, the 309 amino acid residue Olfactory Receptor-like protein from *Rattus norvegicus* (SPTREMBL-ACC:Q63395). In all BLAST herein, the "E-value" or "Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. In addition, it was found, for example, that the GPCR6 nucleic acid sequence has 778 of 927 bases (83%) identical to a gb:GENBANK-ID:AF282289| acc:AF282289.1 mRNA from Mus musculus (Mus musculus odorant receptor K40 gene, complete cds). The full amino acid sequence of the

protein of the invention was found to have 259 of 304 amino acid residues (85%) identical to, and 275 of 304 amino acid residues (90%) similar to, the 308 amino acid residue ptnr:TREMBLNEW-ACC:AAG39874 protein from Mus musculus (Mouse) (ODORANT RECEPTOR K40). Additional BLAST results are shown in Table 6E.

Table 6E. BLAST results for GPCR6					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|7242165|ref|NP_0 35113.1|; gi|3983437|gb|AAD13 307.1| (AF106007); gi|12007413|gb|AAG4 5187.1| (AF321233)	olfactory receptor 41 [Mus musculus]; olfactory receptor I7 [Mus musculus]	327	143/301 (47%)	193/301 (63%)	4e-67
gi|129091|sp|P23267 |; gi|112091|pir||C237 01; gi|205818|gb|AAA417 41.1| (M64378)	OLF6_RAT OLFACTORY RECEPTOR-LIKE PROTEIN F6 [Rattus norvegicus]	311	142/294 (48%)	202/294 (68%)	5e-67
gi|12007417|gb|AAG4 5190.1| (AF321234)	m50 olfactory receptor [Mus musculus]	316	144/303 (47%)	200/303 (65%)	1e-65
gi|12007431|gb|AAG4 5202.1|AF321236_1 (AF321236)	m50 olfactory receptor [Mus musculus]	316	143/303 (47%)	200/303 (65%)	2e-65
gi|6754932|ref|NP_0 35121.1|; gi|3983374|gb|AAD13 315.1| (AF102523)	olfactory receptor 49 [Mus musculus]; olfactory receptor C6 [Mus musculus]	313	142/302 (47%)	196/302 (64%)	4e-65

A multiple sequence alignment is given in Table 6F, with the GPCR6 protein of the invention being shown on line 1, in a ClustalW analysis comparing GPCR6 with related protein sequences.

Table 6F. Information for the ClustalW proteins:

1. GPCR6; SEQ ID NO:18
2. gi\|7242165\|ref\|NP_035113.1\| olfactory receptor 41 [Mus musculus]; SEQ ID NO:55
3. gi\|129091\|sp\|P23267\|OLF6_ rat olfactory receptor - Like Protein F6; SEQ ID NO:56
4. gi\|12007417\|gb\|AAG45190.1\| (AF321234) m50 olfactory receptor [Mus musculus]; SEQ ID NO:57
5. gi\|12007431\|gb\|AAG45202.1\|AF321236_1 m50 olfactory receptor [Mus musculus]; SEQ ID NO:58
6. gi\|6754932\|ref\|NP_035121.1\| olfactory receptor 49 [Mus musculus]; SEQ ID NO:59

```
                      10        20        30        40        50
            ....|....|....|....|....|....|....|....|....|....|
GPCR6       ---MR--NGTVITEFILLG----FPVIQGLQTPLFIAIFLI---------
gi|7242165| MERRN--HTGRVSEEVLLG----FPAPAPLRALLLFLSLLA---------
gi|124091|  MADAFDPSDTRPDELAAVARQKLLVVLGQLQTYIEQVELLKRCDPQVARH
gi|12007417|---ME--NITNISEFIIMG----FPTAPWLQILLLFSIFFLI--------
gi|12007431|---ME--NITNISEFIIMG----FPTAPWLQILLLFSIFFLI--------
gi|6754932| ---MA--NSTTVTEFILLG----LSDACELQVLIELGFLLI--------

                      60        70        80        90        100
            ....|....|....|....|....|....|....|....|....|....|
GPCR6       YILTLAGNGLIIATV---WAEPR------LQIPMYFFLCN---LS-----
```

```
     gi|7242165|   YVLVLTENILIITAT---RNHPT-----LHKPMYFFLAN---MS-----
     gi|124091|    QIGKLKLNALQVRAVSRHFMEGMSSQAATITPLTLALELSLEYARREGE
     gi|12007417|  YVFVLLENLVIILTV---WVTGS-----LHKPMYFLST---MS-----
     gi|12007431|  YVFVLLENLVIILTV---WVTGS-----LHKPMYFLST---MS-----
  5  gi|6754932|   YFLILLGNFLIIFIT---LVDRR-----LYTPMYFLRN---FA-----

                       110       120       130       140       150
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR6         -FLE-----------IWYTTTV---IPKLLGTFVVART----VICMS
 10  gi|7242165|   -FLE-----------IWYVLVT---IPKMLAGFIGSEENHGQLISFE
     gi|124091|    KLLEALNDLGERSSPVAYFEGTMGLARGCPHHQAVKLATYGG---EIDKE
     gi|12007417|  -FLE-----------AWYISVT---VPKMLAGFLFHPN----TISFL
     gi|12007431|  -FLE-----------AWYISVT---VPKMLAGFLFRPN----TISFL
     gi|6754932|   -MLE-----------IWFTSVI---FPKMLTNILTGHK----TISLL
 15

                       160       170       180       190       200
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR6         CCLL-----QA-FFHFFV-GTEFLILTIMSF-DR----YLT---
     gi|7242165|   ACMT-----QL-YFFLGL--GCTECVLLAVMAY-DR----YVA---
 20  gi|124091|    LCPLHDVENFLKQMNYCHLITPASAAAEALVSVKAFLARTVGSELLVPPE
     gi|12007417|  GCMT-----QL-YFFMSL--ACTECVLLAAMAY-DR----YVA---
     gi|12007431|  GCMT-----QL-YFFMSL--ACTECVLLAAMAY-DR----YVA---
     gi|6754932|   GCFL-----QA-FLYFFL--GTTEFFLLAVMSF-DR----YVA---

                       210       220       230       240       250
 25                ....|....|....|....|....|....|....|....|....|....|
     GPCR6         ICNPLHEPTIMTSKLCL---QLALSS-WVVGFTIVFC----QTMI----L
     gi|7242165|   ICHPLHYPVLVSSRLCV---QMAAGS-WAGGFGISMV----KVFL----I
     gi|124091|    ISDPSHPCHVCFEELCVTANQGATASRRLACKICDHVTQQARVRLDADEM
 30  gi|12007417|  ICWPLRYPVMMTTGFCV---QLTISS-WVSGFTISMA----KVYF----I
     gi|12007431|  ICWPLRYPVMMTTGFCV---QLTISS-WVSGFTISMA----KVYF----I
     gi|6754932|   ICNPLRYATIMSKRVCV---QLVFCS-WMSCLLLIEV----PSSI----M

                       260       270       280       290       300
 35                ....|....|....|....|....|....|....|....|....|....|
     GPCR6         IQ-LP-FCG------------------NN---------
     gi|7242165|   SR-LS-YCG------------------PN---------
     gi|124091|    RRNLPHVVGLSEARRARALHALEVSSKMTEANSGGPAEAPGPAAAQEREA
 40  gi|12007417|  SR-VA-FCG------------------NN---------
     gi|12007431|  SR-VA-FCG------------------NN---------
     gi|6754932|   FQ-QP-FCG------------------PN---------

                       310       320       330       340       350
                   ....|....|....|....|....|....|....|....|....|....|
 45  GPCR6         --VIS--HFYCDVGP-SLKAACI--------D-TS---------
     gi|7242165|   --TIN--HFFCDVSP-LLNLSCT--------D-MS---------
     gi|124091|    SALLDAHEVFKSAPPGLYAVSELRFWLSSGDRTSGSTVDAFADNLSALAE
     gi|12007417|  --VLN--HFFCDVSP-ILKLACM--------N-LS---------
     gi|12007431|  --VLN--HFFCDVSP-ILKLACM--------N-LS---------
 50  gi|6754932|   --IIN--HFFCDNFP-LMELICA--------D-TS---------

                       360       370       380       390       400
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR6         ----------ILELLGV------------------IATILV---LPGS
 55  gi|7242165|   ----------TAELTDF------------------IIAIFI---LLGP
     gi|124091|    RERRYETGAVAVELAAFGRRGEHFDRTFGDRVASLDMVDALFVGGQSAAP
     gi|12007417|  ----------MAETVDF------------------ALAIVI---LIFP
     gi|12007431|  ----------MAETVDF------------------ALAIVI---LIFP
     gi|6754932|   ----------LVEFLGF------------------VIANFS---LLGT
 60
                       410       420       430       440       450
```

```
            ....|....|....|....|....|....|....|....|....|....|
  GPCR6        ----LLFNMISYIYILSA-IL--------------------------------
  gi|7242165|  ----LSVTGASYMAITGA-VM--------------------------------
  gi|124091|   DDQIEALVRACYNHHLSAPVLRQLAGSEHGDAEALRSALEGLHAAEDPPG
  gi|12007417| ----LSATVLSYGFIVST-VL--------------------------------
  gi|12007431| ----LSATVLSYGFIVST-VL--------------------------------
  gi|6754932|  ----LAVTATCYGHILYT-IL--------------------------------

                      460       470       480       490       500
            ....|....|....|....|....|....|....|....|....|....|
  GPCR6        ---------RIPSATG----------------------HOKTFS---T-----
  gi|7242165|  ---------RIPSAAG----------------------RHKAFS---T-----
  gi|124091|   DGNAEKEARRAPSLGGGPEDDWAALAARAAADVGARRRLYADRLTKRSLA
  gi|12007417| ---------QIPSATG----------------------QRKAFS---T-----
  gi|12007431| ---------QIPSATG----------------------QRKAFS---T-----
  gi|6754932|  ---------HIPSAKE----------------------RKKAFS---T-----

                      510       520       530       540       550
            ....|....|....|....|....|....|....|....|....|....|
  GPCR6        ----CAS------HLTVVSLLVGAV---LFMYLR-----------P-TA
  gi|7242165|  ----CAS------HLTVVIIFYAAS---IFIYAR-----------P-KA
  gi|124091|   SLGRCVREQRGELEKMLRVSTYGEVLPTVFAAVCNGFAARTRFCELTARA
  gi|12007417| ----CAS------HLTVVVIFYTAV---IFMYVR-----------P-RA
  gi|12007431| ----CAS------HLTVVVIFYTAV---IFMYVR-----------P-RA
  gi|6754932|  ----CSS------HIIVVSLFYGSC---IFMYVR-----------SGKN

                      560       570       580       590       600
            ....|....|....|....|....|....|....|....|....|....|
  GPCR6        HSSFKI--N---------KVVSVLNTILTPLLNPFI----YTLRN-----
  gi|7242165|  LSAFDT--N---------KLVSVLYAVIVPLLNPII----YCLRN-----
  gi|124091|   GTVIDNRGNPDTFDTHRFMRASLMRHRVDPALLPGITHQFFELVNGPLFD
  gi|12007417| IASFNS--N---------KLISAIYAVFTPMLNPII----YCLRN-----
  gi|12007431| IASFNS--N---------KLISAIYAVFTPMLNPII----YCLRN-----
  gi|6754932|  GQGEDH--N---------KVVALLNTVVTPTLNPFI----YTLRN-----

                      610       620       630       640       650
            ....|....|....|....|....|....|....|....|....|....|
  GPCR6        ------KE---------VKGA---------LR---------KAMTCPKT
  gi|7242165|  ------QE---------VKKA---------LR---------RTIHLAQG
  gi|124091|   HATHGFAQPPNTALYFSVENVGLLPHLKEELARFMMGKADSDWAISEFQK
  gi|12007417| ------KE---------VKDA---------LR---------KITAGGRA
  gi|12007431| ------KE---------VKDA---------LR---------KITAGGRA
  gi|6754932|  ------KQ---------VKQV---------FR---------EHVSKFQK

                      660
            ....|....|....|....|..
  GPCR6        G---HAK----------
  gi|7242165|  Q---DANTKKSSRDG--
  gi|124091|   FYHFDGTSGILPTQRIA
  gi|12007417| P---ALGESIS------
  gi|12007431| P---ALGESIS------
  gi|6754932|  F---SQT----------
```

DOMAIN results for GPCR6 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 6G with the statistics and domain description. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to

have significant homology to GPCR6. An alignment of GPCR6 residues 39-288 (SEQ ID NO:18) with 7tm_1 residues 1-254 (SEQ ID NO:34) are shown in Table 6G.

Table 6G. DOMAIN results for GPCR6

PSSMs producing significant alignments:	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	88.2	6e-19

```
                  10        20        30        40        50
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        GNGLIIATVWAEPRLQIPMYFFLCNLSFLEIWYTTIVIPKLLGTFVVA-R
Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGG-D

                  60        70        80        90        100
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        TVICMSCCLLQAFFHFFVGTTEFLILTIMSFDRYLTICNPLHHPTTMISK
Pfam|pfam00001 WVFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPR

                  110       120       130       140       150
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        LCLQIALSSWVVGFTIVFCQTMLLIQLPFCGNNV----ISHFYCDVGPSL
Pfam|pfam00001 RAKVLILLVWVLALLDSLPPLLFSWLRTVEEGNT----TVCLIDFPEESV

                  160       170       180       190       200
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        KAACIDTSILELLGVIATILVIPGSLLFNMISYIYILS------------
Pfam|pfam00001 KRSYVLLSTLVGFVLPLLVILVCYIRILRTLRKRARSQ------------

                  210      - 220       230       240       250
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        --------------------------------------------------
Pfam|pfam00001 --------------------------------------------------

                  260       270       280       290       300
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        ---------------------------------------AILRIPSATGH
Pfam|pfam00001 ---------------------------------------RSLKRRSSSER

                  310       320       330       340       350
             ....|....|....|....|....|....|....|....|....|....|
GPCR6        QKTFSTCASHLTVWSLLYGAVLFMYLRPTAHSSF-----KINKVVSVLNT
Pfam|pfam00001 KAAKMLLVVVVVFVLCWLPYHIVLLLDSLCLLSIW-RVLPTALITLWLA

                  360
             ....|....|.
GPCR6        ILTPLLNPFIY
Pfam|pfam00001 YMNSCLNPIIY
```

The Olfactory Receptor-like GPCR6 proteins disclosed are expressed in at least the following tissues: olfactory receptors are a class of G protein-coupled receptor which are known to be expressed in all tissue types especially olfactory epithelium. Further tissue expression analysis is provided in the Examples.

The nucleic acids and proteins of GPCR6 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further herein.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR6 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR6 epitope is from about amino acids 115 to 135. In additional embodiments, GPCR6 epitopes are from about amino acids 225 to 240 and from about amino acids 280 to 305.

GPCR7

The disclosed novel GPCR7 nucleic acid of 942 nucleotides (also referred to as 20722608_EXT) is shown in Table 7A. An ORF begins with an ATG initiation codon at nucleotides 1-3 and ends with a TGA codon at nucleotides 940-942.

The following genomic clone was identified as having regions of high homology to Olfactory Receptors of the invention: genomic clone >acc:AP000868 HTG *Homo sapiens* chromosome 11 clone RP11-688B18 map 11q24, WORKING DRAFT SEQUENCE, in unordered pieces - *Homo sapiens*, 176200 bp (DNA). The sequence was analyzed by GENSCAN and GRAIL software programs to identify exons and putative coding sequences. The start and stop codons in Table 7A are in bold letters.

Table 7A. GPCR7 Nucleotide Sequence (SEQ ID NO:20)
ATGGGAAACTGGAGCACTGTGACTGAAATCACCCTAATTGCCTTCCCAGCTCTCCTGGAGA
TTCGAATATCTCTCTTCGTGGTTCTTGTGGTAACTTACACATTAACAGCAACAGGAAACAT
CACCATCATCTCCCTGATATGGATTGATCATCGCCTGCAAACTCCAATGTACTTCTTCCTC
AGTAATTTGTCCTTTCTGGATATCTTATACACCACTGTCATTACCCCAAAGTTGTTGGCCT
GCCTCCTAGGAGAAGAGAAAACCATATCTTTTGCTGGTTGCATGATCCAAACATATTTCTA
CTTCTTTCTGGGGACGGTGGAGTTTATCCTCTTGGCGGTGATGTCCTTTGACCGCTACATG
GCTATCTGCGACCCACTGCACTACACGGTCATCATGAACAGCAGGGCCTGCCTTCTGCTGG
TTCTGGGATGCTGGGTGGGAGCCTTCCTGTCTGTGTTGTTTCCAACCATTGTAGTGACAAG
GCTACCTTACTGTAGGAAAGAAATTAATCATTTCTTCTGTGACATTGCCCCTCTTCTTCAG
GTGGCCTGTATAAATACTCACCTCATTGAGAAGATAAACTTTCTCCTCTCTGCCCTTGTCA
TCCTGAGCTCCCTGGCATTCACTACTGGGTCCTACGTGTACATAATTTCTACCATCCTGCG
TATCCCCTCCACCCAGGGCCGTCAGAAAGCTTTTTCTACCTGTGCTTCTCACATCACTGTT
GTCTCCATTGCCCACGGGAGCAACATCTTTGTGTATGTGAGACCCAATCAGAACTCCTCAC
TGGATTATGACAAGGTGGCCGCTGTCCTCATCACAGTGGTGACCCCTCTCCTGAACCCTTT
TATCTACAGCTTGAGGAATGAGAAGGTACAGGAAGTGTTGAGAGAGACAGTGAACAGAATC

ATGACCTTGATACAAAGGAAAACTTGA

The GPCR7 protein (SEQ ID NO:21) encoded by SEQ ID NO:20 has 313 amino acid residues and is presented using the one-letter code in Table 7B. The predicted molecular weight of GPCR7 protein is 35326.06 Daltons. The Psort profile for GPCR7 predicts that this

5   sequence has a signal peptide and is likely to be localized at the endoplasmic reticulum (membrane) with a certainty of 0.685. In alternative embodiments, GPCR7 is located in the plasma membrane with a certainty of 0.676, the Golgi body with a certainty of 0.460 or endoplasmic reticulum (lumen) with a certainty of 0.100. The Signal P predicts a likely cleavage site between positions 36 and 37, i.e., at the dash in the sequence TLT-AT.

10

Table 7B. Encoded GPCR7 protein sequence (SEQ ID NO:21)
MGNWSTVTEITLIAFPALLEIRISLFVVLVVTYTLTATGNITIISLIWIDHRLQTPMYFF LSNLSFLDILYTTVITPKLLACLLGEEKTISFAGCMIQTYFYFFLGTVEFILLAVMSFDR YMAICDPLHYTVIMNSRACLLLVLGCWVGAFLSVLFPTIVVTRLPYCRKEINHFFCDIAP LLQVACINTHLIEKINFLLSALVILSSLAFTTGSYVYIISTILRIPSTQGRQKAFSTCAS HITVVSIAHGSNIFVYVRPNQNSSLDYDKVAAVLITVVTPLLNPFIYSLRNEKVQEVLRE TVNRIMTLIQRKT

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention 20722608_EXT has 581 of 916 bases (63%) identical to a *Mus musculus* Olfactory Receptor-like protein mRNA (GENBANK-ID:AF102523). The full

15   amino acid sequence of the protein of the invention was found to have 149 of 312 amino acid residues (47%) identical to, and 211 of 312 residues (67%) similar to, the 313 amino acid residue Olfactory Receptor-like protein from *Mus musculus* (SPTREMBL-ACC:Q9Z1V0). GPCR7 also has homology to the proteins shown in the BLASTP data in Table 7C.

Table 7C. BLAST results for GPCR7					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|129091|sp|P23267|; gi|112091|pir||C23701; gi|205818|gb|AAA41741. 1| (M64378)	OLF6_RAT OLFACTORY RECEPTOR-LIKE PROTEIN F6 [Rattus norvegicus]	311	141/301 (46%)	184/301 (60%)	1e-63
gi|6754932|ref|NP_0351 21.1|; gi|3983374|gb|AAD13315 .1| (AF102523)	olfactory receptor 49 [Mus musculus]; olfactory receptor C6 [Mus musculus]	313	142/312 (45%)	195/312 (61%)	2e-63

gi|7242165|ref|NP_0351 13.1|; gi|3983437|gb|AAD13307 .1| (AF106007); gi|12007413|gb|AAG4518 7.1| (AF321233)	olfactory receptor 41; olfactory receptor I7 [Mus musculus]	327	133/302 (44%)	181/302 (59%)	5e-62
gi|10181106|ref|NP_065 623.1|; gi|7638409|gb|AAF65461 .1|AF247657_1 (AF247657); gi|12007410|gb|AAG4518 4.1| (AF321233)	olfactory receptor 17; olfactory receptor P2; P2 olfactory receptor [Mus musculus]	315	133/307 (43%)	184/307 (59%)	2e-61
gi|13928994|ref|NP_113 898.1|; gi|129092|sp|P23270|; gi|112099|pir||F2370|; gi|205834|gb|AAA41749. 1| (M64386)	olfactory receptor 41; OLF7_RAT olfactory receptor -like protein I7 olfactory receptor I7 - rat [Rattus norvegicus]	327	131/302 (43%)	179/302 (58%)	2e-60

A multiple sequence alignment is given in Table 7D, with the protein of the invention 20722608_EXT being shown on line 1, in a ClustalW analysis comparing the protein of the invention with related protein sequences. This BLASTP data is displayed graphically in the ClustalW in Table 7D.

Table 7D. ClustalW Analysis of GPCR7

1) GPCR7; SEQ ID NO:21
2) >gi\|129091\|sp\|P23267\|OLF6_rat olfactory receptor - Like Protein F6; SEQ ID NO:60
3) >gi\|6754932\|ref\|NP_035121.1\| olfactory receptor 49 [Mus musculus] ; SEQ ID NO:61
4) >gi\|7242165\|ref\|NP_035113.1\| olfactory receptor 41 [Mus musculus] ; SEQ ID NO:62
5) >gi\|10181106\|ref\|NP_065623.1\| olfactory receptor 17 [Mus musculus] ; SEQ ID NO:63
6) >gi\|13928994\|ref\|NP_113898.1\| olfactory receptor 41 [Rattus norvegicus] ; SEQ ID NO:64

```
                      10        20        30        40        50
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         -----MGNWSTVTEITLIAFP-ALLEIRISLFVVLVVTYTLTATGNITII
gi|129091|    MAWSTGQNLSTPGPFILLGFP-GPRSMRIGLFLLFLVMYLLTVVGNLAII
gi|6754932|   -----MANSTTVTEFILLGLS-DACELQVLIFLGFLLTYFLILLGNFLII
gi|7242165|   --MERRNHTGRVSEFVLLGFP-APAPLRALLFFLSLLAYVLVLTENILII
gi|10181106|  ---MTWGNWTTVREFILMSFSSLSYEVQALLFLLFLIIYLVTLMGNVLII
gi|13928994|  --MERRNHSGRVSEFVLLGFP-APAPLRVLLFFLSLLXYVLVLTENMLII

                      60        70        80        90        100
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         SLIWIDHRLQTPMYFFLSNLSFLDILYTTVITPKLLACLLG----EEKTI
gi|129091|    SLVGAHRCLQTPMYFFLCNLSFLEIWFTTACVPKTLATFAP----RGGVI
gi|6754932|   FITLVDRRLYTPMYYFLRNFAMLEIWFTSVIFPKMLTNIIT----GHKTI
gi|7242165|   TAIRNHPTLHKPMYFFLANMSFLEIWYVTVTIPKMLAGFIGSEENHGQLI
gi|10181106|  LVTTADSALQSPMYFFLRNLSFLEIGFNLVIVPKMLSTLIL----QDKTI
gi|13928994|  IAIRNHPTLHKPMYFFLANMSFLEIWYVTVTIPKMLAGFIGSKENHGQLI

                      110       120       130       140       150
              ....|....|....|....|....|....|....|....|....|....|
```

```
GPCR7            SFAGCMIQTYFYFFLGTVEFILLAVMSFDRYMAICDPLHYTVIMNSRACL
gi|129091|       SLAGCATQMYEVFSLGCTEYFLLAVMAYDRYLAICLPLRYGGIMTPGLAM
gi|6754932|      SLLGCFLQAFLYFFLGTTEFFLLAVMSFDRYVAICNPLRYATIMSKRVCV
gi|7242165|      SFEACMTQLYFFLGLGCTECVLLAVMAYDRYVAICHPLHYPVIVSSRLCV
gi|10181106|     SFLGCATQMYFFFFGAAECCLLATMAYDRYMAICDPLHYPIIMSRRSCA
gi|13928994|     SFEACMTQLYFFLGLGCTECVLLAVMAYDRYVAICHPLHYPVIVSSRLCV
```

```
                        160       170       180       190       200
                 ....|....|....|....|....|....|....|....|....|....|
GPCR7            LLVLGCWVGAFLSVLFPTIVVTRLPYCR-KEINHFFCDIAPLLQVACINT
gi|129091|       RLALGSWLCGFSAITVPATLIARLSFCGSRVINHFFCDISPWIVLSCTDT
gi|6754932|      QLVFCSWMSGLLLIIVPSSIVFQQPFCGPNIINHFFCDNFPIMELICADT
gi|7242165|      QMAAGSWAGGFGISMVKVFLISRLSYCGPNTINHFFCDVSPLLNLSCTDM
gi|10181106|     QLAAASWFSGFPVATVQTTWIFSFPFCGPNMVNHFFCDSPPVIALVCADT
gi|13928994|     QMAAGSWAGGFGISMVKVFLISRLSYCGPNTINHFFCDVSPLLNLSCTDM
```

```
                        210       220       230       240       250
                 ....|....|....|....|....|....|....|....|....|....|
GPCR7            HLIEKINFLLSALVILSSLAFTTGSYVYIISTILRIPSTQGROKAFSTCA
gi|129091|       QVVELVSFGIAFCVILGSCGITLVSYAYIITTIIKIPSARGRHRAFSTCS
gi|6754932|      SLVEFLGFVIANFSLLGTLAVTATCYGHILYTILHIPSAKERKKAFSTCS
gi|7242165|      STAELTDFILAIFILLGPLSVTGASYMAITGAVMRIPSAAGRHKAFSTCA
gi|10181106|     SLFELEALTATVLFILFPFLLILGSYVRILSTIFRMPSAEGKRKAFSTCS
gi|13928994|     STAELTDFVLAIFILLGPLSVTGASYMAITGAVMRIPSAAGRHKAFSTCA
```

```
                        260       270       280       290       300
                 ....|....|....|....|....|....|....|....|....|....|
GPCR7            SHITVVSIAHGSNIFVYVRPNQNSSLD-YDKVAAVLITVVTPLLNPFIYS
gi|129091|       SHLTVVLIWYGSTIFLHVRTSVESSLD-LTKAITVLNTIVTPVLNPFIYT
gi|6754932|      SHIIVVSLFYGSCIFMYVRSGKNGQGEDHNKVVALLNTVVTPTLNPFIYT
gi|7242165|      SHLTVVIIFYAASIFIYARPKALSAFD-TNKLVSVLYAVIVPLLNPIIYC
gi|10181106|     SHLLVVSLFYSTAILTYFRPRSNTSPE-NKKMLSLSYTVITPMLNPIIYS
gi|13928994|     SHLTVVIIFYAASIFIYARPKALSAFD-TNKLVSVLYAVIVPLFNPIIYC
```

```
                        310       320       330       340       350
                 ....|....|....|....|....|....|....|....|....|....|
GPCR7            LRNEKVQEVLRETVNRIMTLIQRKT-------------------------
gi|129091|       LRNKDVKEALRRIVKGK---------------------------------
gi|6754932|      LRNKQVKQVFREHVSKFQKFSQT---------------------------
gi|7242165|      LRNQEVKKALRRTLHLAQGQDANTKKSSRDG-------------------
gi|10181106|     LRNNEVKAALRRIIHRTLGPQKL---------------------------
gi|13928994|     LRNQDVKRALRRTLHLAQDQEANTNKGSKIGUNABLETFINDTHESEQUE
```

```
                 ....|....
GPCR7            ---------
gi|129091|       ---------
gi|6754932|      ---------
gi|7242165|      ---------
gi|10181106|     ---------
gi|13928994|     NCEWITHGI
```

Table 7E lists the domain description from DOMAIN analysis results against GPCR7.

This indicates that the GPCR7 sequence has properties similar to those of other proteins

49

known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID NO:34) itself.

Table 7E Domain Analysis of GPCR7

```
PSSMs producing significant alignments:                        Score  E
                                                               (bits) value
gnl|Pfam|pfam00001  7tm_1, 7 transmembrane receptor (rhodopsin family) 108  4e-25


                       10        20        30        40        50
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         GNITIISLIWIDHRLQ--TPMYFFLSNLSFLDILYTTVLTPKLLACLLGE
Pfam|pfam00001 GNLLVILVILRTKKLR--TPTNIFLLNLAVADLLRLLTLPPWALYYLVGG


                       60        70        80        90       100
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         -EKTISFAGCMIQTYFYFFLGTVEFILLAVMSFDRYMAICDPLHYTVIMN
Pfam|pfam00001 -DWVFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRT


                      110       120       130       140       150
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         SRACLLLVLGCWVGAFLSVLFPTIVVTRLPYCRKEINH-----FFCDIAP
Pfam|pfam00001 PRRAKVLILLVWVLALLLSLPPLLFSWLRTV--EEGNT-----TVCLIDF


                      160       170       180       190       200
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         LLQVAC------INTHLIEKLNFLLSALVILSSLAFTTG---SYVYIIS
Pfam|pfam00001 PEESVK------RSYVLLSTLVGFVLPLLVILVCYTRILRTLRKRARSQ


                      210       220       230       240       250
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         ------------------------------------------------
Pfam|pfam00001 ------------------------------------------------


                      260       270       280       290       300
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         ------------------------------------------------
Pfam|pfam00001 ------------------------------------------------


                      310       320       330       340       350
              ....|....|....|....|....|....|....|....|....|....|
GPCR7         -TILRIPSTQGRQKAFSTCASHLTVVSIAHGSNIFVYVRPNQNSS-----
Pfam|pfam00001 -RSLKRRSSSERKAAKMLLVVVVVFVLCWLPYHIVLLLDSLCLLSIW-RV


                      360       370
              ....|....|....|....|....|
GPCR7         LDYDKVAAVLITVVTPLLNPFIY
Pfam|pfam00001 LPTALLITLWLAYVNSCLNPIIY
```

The GPCR7 protein disclosed in this invention is expressed in at least the following human tissues: pancreas and olfactory epithelium; This is by no way limiting in that olfactory receptors are a class of G protein-coupled receptor which are known to be expressed in all tissue types.

The nucleic acids and proteins of GPCR7 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further herein.

The novel GPCR7 nucleic acid encoding the GPCR-like protein of the invention, or

5   fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the

10   art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR7 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR7 epitope is from about amino acids 120 to 132. In additional embodiments, GPCR7 epitopes are from about amino acids 160 to 180, from about amino acids 215 to 240, from about amino

15   acids 255 to 270 and from about amino acids 280 to 310.

GPCR8

The disclosed novel GPCR8 nucleic acid of 1920 nucleotides (also referred to as CG-SC931712) is shown in Table 8A. An ORF begins with an ATG initiation codon at nucleotides 201-203 and ends with a TGA codon at nucleotides 1137-1139.

20   The following genomic clone was identified as having high homology to olfactory receptor-like protein (HS6M1-6). The start and stop codons in Table 8A are in bold letters and the putative untranslated regions upstream from the initiation site and downstream from the termination codon are underlined.

Table 8A. GPCR8 Nucleotide Sequence (SEQ ID NO:22)
ATTGCTTGTATCTGTAGCTGGATAAATATCTCAATGAAGCATATAAAGGGAACTGTATAAAAATTC
TACTACCATTATGGTGCACACTCTCTGGAAGTGGGATACTTTTGTCTTCAATCTGTTTGCAAGTGA
GCGGTTGACAATGCATGGACAGACTTTGAGTTTATGTGGTTCTTTCTTTAGGTATAAGAAAAAGAT
GAATGATGATTAAAAAAAATGCAAGTTCGGAAGACTTCTTTATTCTACTTGGATTTTCTAATTGGC
CTCAGCTGGAAGTAGTTCTCTTTGTGGTTATCTTGATCTTCTACCTGATGACACTGACAGGAAACC
TGTTCATCATCATCCTGTCATACGTGGACTCCCATCTCCACACACCAATGTACTTCTTCCTTTCAA
ACCTCTCATTTCTGGATCTCTGCCACACCACCAGCTCTATCCCTCAGTTGCTGGTGAATCTCCGGG
GCCCGGAAAAGACCATCTCGTATGCTGGTTGCATGGTTCAACTTTACTTTGTTCTTGCACTGGGAA
TCGCAGAGTGTGTCCTACTGGTGGTGATGTCGTATGATCGTTATGTAGCTGTGTGTAGACCTTTGC
ATTACACTGTCCTCATGCACCCTCGTTTCTGCCACTTGTTGGCTGCGGCTTCTTGGGTAATTGGTT
TTACTATCTCAGCACTTCATTCCTCCTTTACTTTCTGGGTACCCCTTTGTGGACATCGCCTAGTGG
ATCACTTCTTCTGTGAAGTTCCAGCACTTCTGCGTTTATCATGTGTTGACACCCATGCAAATGAGC

```
TGACCCTCATGGTCATGAGCTCCATTTTTGTTCTCATACCTCTCATTCTGATTCTCACTGCCTATG
GTGCCATTGCCCGGGCTGTACTGAGCATGCAATCAACCACTGGGCTTCAGAAAGTGTTTAGGACAT
GTGGAGCCCATCTTATGGTTGTATCTCTCTTTTTCATTCCAGTCATGTGCATGTATCTCCAGCCAC
CATCAGAAAATTCTCCTGATCAGGGCAAGTTCATTGCCCTCTTTTATACTGTTGTCACACCGAGTC
TTAATCCTCTAATCTACACTCTCAGAAACAAGCATGTAAAAGGGGCAGCGAAGAGACTATTGGGGT
GGGAGTGGGGGAAGTGACAGGGAAATCATGTTGTCTGTTGTCATTGTTTTTCCTAGGGTCTTAGCC
ATCTTGAAAGGTGGTTTCCCTGCTTCTTTGTGATTTATTTTTGTTCTAACAGCTCACAAAACATGG
AATAGTTCAGTTCCCCCATTTGTTGCTCTGTTTAATATTTAGTTCTGAAATATTATGTTGAGATAA
AGGTTTTGATTAGTACCATTTTGTTCTTTTACAATTGTATATTTTATTTCTGTGAAAATTGTGGAC
TGTGGTTTCAACGTAAATAAATGTGCATGCGAATAGTTATGAGGAGATTATTTAAAAAATATTGGC
AATATTTCTGACAATGTGCTAAATTATGAACTGACCATTGATATGTATAGGAAGAGAAGGGCAATA
TTGCAAAGATGTAGGCTGAAGAAGTTTTTGGTTATTAAATAAACCTTAAATGAAGCTAAAAATAGT
CACAGCAAAGAAAAATAGTAAACATAATGAATAACACCATTTATTATATGGTAAAGGATATGTCAT
AATTTTTTGGTTGAAGTTCACTTTTTAAAGACACTAAATTATATAATTTATCCTGTAGGTCTGCAT
TCTTGTCACATTGAACAGTAAACTAATATCTCTTTAAAATGGCTGATTCGTTCATCTGTCCATTTA
TTCATTAACTTATTCTTCATTAGCTAAATCTTACTGGACATGTACTCTCTCCCAGTTTGTGAAATT
CTTGGTAACATGTATAAATATAACATACTTTGTCTGAACAGAATGCACTCTCTATCGGGAAAAATG
GCAACA
```

The GPCR8 protein (SEQ ID NO:23) encoded by SEQ ID NO:22 has 312 amino acid residues and is presented using the one-letter code in Table 8B. The predicted molecular weight of GPCR8 protein is 35202.21 Daltons. The GPCR8 amino acid sequence is 100% homologous to olfactory receptor-like protein (HS6M1-6). The Psort profile for GPCR8 predicts that this sequence has a signal peptide and is likely to be localized at the endoplasmic reticulum (membrane). In an alternative embodiment, GPCR8 is located in the plasma membrane.

Table 8B. Encoded GPCR8 protein sequence (SEQ ID NO:23)
MMIKKNASSEDFFILLGFSNWPQLEVVLFVVILIFYLMTLTGNLFIIILSYVDSHLHTPMYFFLSNLS FLDLCHTTSSIPQLLVNLRGPEKTISYAGCMVQLYFVLALGIAECVLLVVMSYDRYVAVCRPLHYTVL MHPRFCHLLAAASWVIGFTISALHSSFTFWVPLCGHRLVDHFFCEVPALLRLSCVDTHANELTLMVMS SIFVLIPLILILTAYGAIARAVLSMQSTTGLQKVFRTCGAHLMVVSLFFIPVMCMYLQPPSENSPDQG KFIALFYTVVTPSLNPLIYTLRNKHVKGAAKRLLGWEWGK

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention CG-SC931712 is greater than 95% homologous to olfactory receptor-like protein (HS6M1-6) and gi|13624331 GPCR8 also has homology to the proteins shown in the BLASTP data in Table 8C.

Table 8C. BLAST results for GPCR8					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect

gi|13624331|ref|NP_112 167.1| ; gi|14423776|sp|O76002| ; gi|3757728|emb|CAA1878 4.1| (AL022727); gi|12054367|emb|CAC204 91.1| (AJ302571)	olfactory receptor, family 2, subfamily J, member 2 [Homo sapiens]; O2J2_HUMAN OLFACTORY RECEPTOR 2J2 (OLFACTORY RECEPTOR 6-8) (OR6-8) (HS6M1-6); dJ80I19.4 (olfactory receptor-like protein (hs6M1-6)); olfactory receptor [Homo sapiens]	312	302/312 (96%)	302/312 (96%)	e-154
gi|12054379|emb|CAC204 97.1| (AJ302577); through gi|12054389|emb|CAC205 02.1| (AJ302582)	olfactory receptor [Homo sapiens]	312	299/312 (95%)	300/312 (95%)	e-152
gi|12054391|emb|CAC205 03.1| (AJ302583)	olfactory receptor [Homo sapiens]	312	298/312 (95%)	299/312 (95%)	e-151
gi|12054355|emb|CAC204 85.1| (AJ302565); through gi|12054365|emb|CAC204 90.1| (AJ302570)	olfactory receptor [Homo sapiens]	312	275/311 (88%)	284/311 (90%)	e-139
gi|12054359|emb|CAC204 87.1| (AJ302567); gi|12054361|emb|CAC204 88.1| (AJ302568)	olfactory receptor [Homo sapiens]	312	274/311 (88%)	284/311 (91%)	e-139

A multiple sequence alignment is given in Table 8D, with the protein of the invention 20722608_EXT being shown on line 1, in a ClustalW analysis comparing the protein of the invention with related protein sequences, shown in Table 8C. This BLASTP data is displayed graphically in the Clustal W in Table 8D.

Table 8D. ClustalW Analysis of GPCR8

1) GPCR8; SEQ ID NO:23
2) >gi\|13624331\|ref\|NP_112167.1\| olfactory receptor, family 2, subfamily J, member 2 [Homo sapiens]; SEQ ID NO:65
3) >gi\|12054379\|emb\|CAC20497.1\| AJ302577 olfactory receptor [Homo sapiens]; SEQ ID NO:66
4) >gi\|12054391\|emb\|CAC20503.1\| AJ302583 olfactory receptor [Homo sapiens]; SEQ ID NO:67
5) >gi\|12054355\|emb\|CAC20485.1\| AJ302565 olfactory receptor [Homo sapiens]; SEQ ID NO:68
6) >gi\|12054359\|emb\|CAC20487.1\| AJ302567 olfactory receptor [Homo sapiens]; SEQ ID NO:69

```
                    10        20        30        40        50
               ....|....|....|....|....|....|....|....|....|....|
GPCR8          MMIKKNASSEDFFILLGFSNWPQLEVVLFVVILIFYLMTLTGNLFIIILS
gi|13624331|   MMIKKNASSEDFFILLGFSNWPQLEVVLFVVILIFYLMTLTGNLFIIILS
gi|12054379|   MMIKKNASSEDFFILLGFSNWPQLEVVLFVVILIFYLMTLTGNLFIIILS
gi|12054391|   MMIKKNASSEDFFILLGFSNWPQLEVVLFVVILIFYLMTLTGNLFIIILS
gi|12054355|   MLMKKNASEEDFFILLGFSNWPHLEVVLFVVILIFYLITLTGNLFIIILS
gi|12054359|   MLMKKNASEEDFFLLLGFSNWPHLEVVLFVVILIFYLITLTGNLFIIILS

                    60        70        80        90        100
               ....|....|....|....|....|....|....|....|....|....|
GPCR8          YVDSHLHTPMYFFLSNLSFLDLCHTTSSIPQLLVNLRGPEKTISYAGCMV
gi|13624331|   YVDSHLHTPMYFFLSNLSFLDLCHTTSSIPQLLVNLRGPEKTISYAGCMV
gi|12054379|   YVDSHLHTPMYFFLSNLSFLDLCYTTSSIPQLLVNLRGPEKTISYAGCMV
```

```
         gi|12054391|  YVDSHLHTPMYFFLSNLSFLDLCYTTSSIPQLLVNLRGPEKTISYAGCMV
         gi|12054355|  YLDSHLHTPMYFFLSNLSFLDLCYTTSSIPQLLVNLWGPEKTISYAGCTV
         gi|12054359|  YLDSHLHTPMYFFLSNLSFLDLCYTTSSIPQLLVNLWGPEKTISYAGCTV
```

5

```
                          110       120       130       140       150
                      ....|....|....|....|....|....|....|....|....|....|
         GPCR8         QLYFVLALGIAECVLLVVMSYDRYVAVCRPLHYTVLMHPRFCHLLAAASW
         gi|13624331|  QLYFVLALGIAECVLLVVMSYDRYVAVCRPLHYTVLMHPRFCHLLAAASW
         gi|12054379|  QLYFVLALGIAECVLLVVMSYDRYVAVCRPLHYTVLMHPRFCHLLVAASW
10       gi|12054391|  QLYFVLALGITECVLLVVMSYDRYVAVCRPLHYTVLMHPRFCHLLVAASW
         gi|12054355|  QLYFVLALGTAECVLLVVMSYDRYAAVCRPLHYTVLMHPRFCRLLAAASW
         gi|12054359|  QLYFVLALGTAECVLLVVMSYDRYAAVCRPLHYTVLMHPRFCRLLAAASW
```

15

```
                          160       170       180       190       200
                      ....|....|....|....|....|....|....|....|....|....|
         GPCR8         VIGFTISALHSSFTFWVPLCGHRLVDHFFCEVPALLRLSCVDTHANELTL
         gi|13624331|  VIGFTISALHSSFTFWVPLCGHRLVDHFFCEVPALLRLSCVDTHANELTL
         gi|12054379|  VIGFTISALHSSFTFWVPLCGHRLVDHFFCEVPALLRLSCVDTHANELTL
         gi|12054391|  VIGFTISALHSSFTFWVPLCGHRLVDHFFCEVPALLRLSCVDTHANELTL
20       gi|12054355|  VSGFTTSALHSSFTFWIPLCRHRLVDHFFCEVPALLRLSCVDTQANELTL
         gi|12054359|  VSGFTTSALHSSFTFWIPLCRHRLVDHFFCEVPALLRLSCVDTQANELTL
```

```
                          210      220       230       240       250
                      ....|....|....|....|....|....|....|....|....|....
25 ·     GPCR8         MVMSSIFVLIPLILILTAYGAIARAVLSMQSTTGLQKVFRTCGAHLMVVS
         gi|13624331|  MVMSSIFVLIPLILILTAYGAIARAVLSMQSTTGLQKVFRTCGAHLMVVS
         gi|12054379|  MVMSSIFVLIPLILILTTYGAIARAVLSMQSTTGLQKVFRTCGAHLMVVS
         gi|12054391|  MVMSSIFVLIPLILILTTYGAIARAVLSMQSTTGLQKVFRTCGAHLMVVS
         gi|12054355|  MVMSSIFVLIPLILILTSYGAIARAVLSMQSTTGLQKVLRTCGAHLMVVS
30       gi|12054359|  MVMSSIFVLIPLILILTSYGAIARAVLSMQSTTGLQKVLRTCGAHLMVVS
```

```
                          260       270       280       290       300
                      ....|....|....|....|....|....|....|....|....|....|
         GPCR8         LFFIPVMCMYLQPPSENSPDQGKFIALFYTVVTPSLNPLIYTLRNKHVKG
35       gi|13624331|  LFFIPVMCMYLQPPSENSPDQGKFIALFYTVVTPSLNPLIYTLRNKHVKG
         gi|12054379|  LFFIPVMCMYLQPPSENSPDQGKFIALFYTVVTPSLNPLIYTLRNKHVKG
         gi|12054391|  LFFIPVMCMYLQPPSENSPDQGKFIALFYTVVTPSLNPLIYTLRNKHVKG
         gi|12054355|  LFFIPVMCMYLQPPSENSQDQGKFIALFYTVVTPSLNPLIYTFRNKDVRG
         gi|12054359|  LFFIPVMCMYLQPPSENSQDQGKFIALFYTVVTPSLNPLIYTFRNKDVRG
```

40

```
                          310
                      ....|....|..
         GPCR8         AAKRLLGWEWGK
         gi|13624331|  AAKRLLGWEWGK
45       gi|12054379|  AAKRLLGWEWGK
         gi|12054391|  AAKRLLGWEWGK
         gi|12054355|  AVKRLMGWEWGM
         gi|12054359|  AVKRLMGWEWGM
```

50

Table 8E lists the domain description from DOMAIN analysis results against GPCR8.
This indicates that the GPCR8 sequence has properties similar to those of other proteins
known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID NO:34)
itself.

55

Table 8E Domain Analysis of GPCR8

PSSMs producing significant alignments: Score E

```
                                                          (bits) value
gnl|Pfam|pfam00001   7tm_1, 7 transmembrane receptor (rhodopsin family) 111    7e-26


                          10        20        30        40        50
                  ....|....|....|....|....|....|....|....|....|....|
 5
       GPCR8      GNLFIIILSYVDSHLHIPMYFFLSNLSFLDLCHTTSSIPQLLVNLRCPEK
       Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDW


                          60        70        80        90        100
                  ....|....|....|....|....|....|....|....|....|....|
10
       GPCR8      TISYAGCMVQLYFVLALGIAECVLLVVMSYDRYVAVCRPLHYTVLMHPRF
       Pfam|pfam00001 VFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRRIRTPRR


                          110       120       130       140       150
                  ....|....|....|....|....|....|....|....|....|....|
15
       GPCR8      CHLLAAASWVIGFTISALHSSFTFWVPLCGHRLVDHFFCEVP----AFLR
       Pfam|pfam00001 AKVLILLVWVLALLLSLPPLLF--------SWLRTVEEGNT----TVCL


                          160       170       180       190       200
                  ....|....|....|....|....|....|....|....|....|....|
20
       GPCR8      LSCVDTHANELTLMVMSSFFVLTPIILLTAVGAIARAVLS---------
       Pfam|pfam00001 IDFPEESVKRSYVLLSTLVGFVLPLLVIIVCYTRILRTLRKRARSQ----


                          210       220       230       240       250
                  ....|....|....|....|....|....|....|....|....|....|
25
       GPCR8      --------------------------------------------------
       Pfam|pfam00001 --------------------------------------------------


                          260       270       280       290       300
                  ....|....|....|....|....|....|....|....|....|....|
30
       GPCR8      --------------------------------------------------
       Pfam|pfam00001 --------------------------------------------------


                          310       320       330       340       350
                  ....|....|....|....|....|....|....|....|....|....|
35
       GPCR8      ----------------------MQSTTGLQKVFRTCGAHLMVVSL
       Pfam|pfam00001 ----------------------RSLKKRSSSERKAAKMLLVVAVVVFVL


                          360       370       380       390
                  ....|....|....|....|....|....|....|....|....|.
40
       GPCR8      FFIP-----VMCMYLQPPSE-NSPDQGKFIALFYTVVTPSLNPLIY
       Pfam|pfam00001 CWLPYHIVLLLDSLCLLSIW-RVLPTALLITLWLAYVNSCLNPLIY
```

45   The Olfactory Receptor-like protein disclosed in this invention is expressed in at least

the following human tissues: pancreas and olfactory epithelium; This is by no way limiting in

that olfactory receptors are a class of G protein-coupled receptor which are known to be

expressed in all tissue types.

          The GPCR8 nucleic acids and proteins are useful in potential therapeutic applications

50   implicated in various GPCR-related pathological disorders and/or OR-related pathological

disorders, described further herein. A monoclonal antibody targeting CG-SC931712 protein,

specifically its extracellular region, will have a therapeutic role in treating cancer. It will also

have a role in treating angiogenesis related diseases. Being a GPCR, it could be used to screen for small molecule drug to treat cancer.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or

5   diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR8 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR8 epitope is from about amino acids 1 to 15. In additional embodiments, GPCR8 epitopes are

10   from about amino acids 80 to 95, from about amino acids 115 to 130, from about amino acids 165 to 175, from about amino acids 180 to 195, from about amino acids 230 to 245, from about amino acids 255 to 270 and from about amino acids 285 to 305.

GPCR9

A second GPCR-like protein of the invention, referred to herein as GPCR9, is an

15   Olfactory Receptor ("OR")-like protein. The GPCR9 gene maps to chromosome 9 p13.1-13.3. Some members of the Olfactory Receptor-Like Protein Family end up localized at the cell surface, where they exhibit activity. Therefore it is likely that these novel GPCR9 proteins are available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

20   Two alternative novel GPCR9 nucleic acids, namely GPCR9a and GPCR9b, and an encoded GPCR9 polypeptide are provided,.

GPCR9a

In one embodiment, a GPCR9 variant is the novel GPCR9a (alternatively referred to herein as 21629632.0.20), which includes the 2028 nucleotide sequence (SEQ ID NO:24)

25   shown in Table 9A. A GPCR9a ORF begins with a Kozak consensus ATG initiation codon at nucleotides 469-471 and ends with a TGA codon at nucleotides 1447-1449. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 9A, and the start and stop codons are in bold letters.

Table 9A. GPCR9 Nucleotide Sequence (SEQ ID NO:24)
TGCTATAGCCCCAGCACTTGATACCTAGCACAGAATAGGTACTTAATAAATACTTAGTGGATGAAT
AAATCTGAAATACTATGGCCATAATTTGGTCACATGAAGCCGTAATGTAGAAAAGATGCTTCCTGT
TAATGACCAAAAACACTTTGGATTCCAAACGATCATTTTAAACATGAATCTTTCTCTGCTGTCTCC
TCTGACCCCATCCTGGGGAGAGCAGAGAGGAGCCTAGGGGACTAGAATGTGCCCCATCCTCCCCTC
AGTGACGTCCACAGAACTGCAGCGCTGAGAAGGCCAGATTGCAGATCTGAAGTCCAACTCCCTCAT

```
TATACAGATGGTGAAACTAAATTCCAGAGAGGGAGGCTGACCTGCTGCAGCTCAGACATCAGGTCA
CTGGGCTCCCAGGCCAGTTGGAGCTTTTTCCAAAAAGCTGGGTGGTCCAGATGGAAAAGGAGAGAG
AATGAGATGAAGTGGGCAAACCAGACAGCTGTGACGGAATACGTCCTGATGGGGCTACACGAGCAC
TGTAACCTGGAGGTGGTCCTGTTTGTGTTCTGCCTGGGCATCTACTCCGTGAATGTGTTGGGGAAC
GCCCTCCTCATAGGGCTGAACGTGCTGCACCCTCGCCTGCACAACCCCATGTACTTCCTTCTCAGC
AACCTCTCCCTCATGGACATCTGCGGCACCTCCTCCTTTGTGCCTCTCATGCTAGACAATTTCCTG
GAAACCCAGAGGACCATTTCCTTCCCTGGCTGTGCCCTGCAGATGTACCTGACCCTGGCGCTGGGA
TCAACGGAGTGCCTGCTGCTGGCTGTGATGGCATATGACCGTTATGTGGCTATCTGCCAGCCGCTT
AGGTACCCAGAGCTCATGAGTGGGCAGACCTGCATGCAGATGGCAGCGCTGAGCTGGGGGACAGGC
TTTGCCAACTCACTGCTACAGTCCATCCTTGTCTGGCACCTCCCCTTCTGTGGCCACGTCATCAAC
TACTTCTATGAGATCTTGGCAGTGCTAAAACTGGCCTGTGGGGACATCTCCCTCAATGCGCTGGCA
TTAATGGTGGCCACAGCCGTCCTGACACTGGCCCCCCTCTTGCTCATCTGCCTGTCTTACCTTTTC
ATCCTGTCTGCCATCCTTAGGGTACCCTCTGCTGCAGGCCGGTGCAAAGCCTTCTCCACCTGCTCA
GCCCACCGCACAGTGGTGGTGGTTTTTTATGGGACAATCTCCTTCATGTACTTCAAACCCAAGGCC
AAGGATCCCAACGTGGATAAGACTGTCGCATTGTTCTACGGGGTTGTGACGCCCTCGCTGAACCCC
ATCATTTACAGCCTGAGGAATGCAGAGGTGAAAGCTGCCGTCCTAACTCTGCTGAGAGGAGGTTTG
CTCTCCAGGAAAGCATCCCACTGCTACTGCTGCCCTCTGCCCCTGTCAGCTGGCATAGGCTAGGTT
GTGCTGTGGTCATGACCTCAAACCTTGAGAGGCTTAAAGCCATTAAGGTTTGTTTCTTGCTCCTGA
TGCAGGTCCACCAGAGGCTGGTGGGGCTTCTGCTCCGCATCATGGTCTTCACCCCTCTGGGACTCA
GGATGACAAAACAGCTACCATTGGGAACACTGCTGGTCACCATGACAAAAAGAAAAGGGAAAGTAA
CAAAGCCTACACTGACTCTTAAAGCTTCTACTCAGAAGTGGCTGTGTTGCCTCCACCTACATTTCA
GTGGCCAACACAATGGCAACAGGAAGGCACAGGACAGAGAGATTTGTGAATGGCCTAATGACTACCACAC
ATCGTGTGTCTGGATGGCAAACGAGAGGGACAGACAGAGATTTGTGAATGGCCTAATGACTACCACAC
CAGCTGACAGTGTCAACCCAAGAGCTATGGGAGGTTTGGCTTTCTTTATCCTGACCATCTATCCTT
CACGGGCTGCTGCCAAGTTAATCGTCCCAAGAAAGCTCTGGTTAGCTCACGTGTGGTAGCTTTATA
CTGAGTCAACCAAACTAGGCTAGAGGGTGTGGGTTAGGGTTGGCCACA
```

The sequence of GPCR9a was derived by laboratory cloning of cDNA fragments, by *in silico* prediction of the sequence. The cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, were cloned. *In silico* prediction was based

5   on sequences available in CuraGen's proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The cDNA coding for the GPCR9a sequence was cloned by the polymerase chain reaction (PCR). Primers were designed based on *in silico* predictions of the full length or some

10   portion (one or more exons) of the cDNA/protein sequence of the invention. The DNA sequence and protein sequence for a novel Olfactory Receptor-like gene were obtained by exon linking and are reported here as GPCR9a.These primers and methods used to amplify GPCR9 a cDNA are described in the Examples.

The GPCR9a polypeptide (SEQ ID NO:25) encoded by SEQ ID NO:24 is 326 aa in

15   length, has a molecular weight of 35713.69 Daltons, and is presented using the one-letter amino acid code in Table 9B. The Psort profile for GPCR9 predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.600. In alternative embodiments, a GPCR9 polypeptide is located to the Golgi body with a certainty of 0.400, the endoplasmic reticulum (membrane) with a certainty of 0.300, or a mitochondrial

inner membrane with a certainty of 0.300. The Signal P software program predicts no likely

signal cleavage site for a GPCR9 peptide.

Table 9B. GPCR9a protein sequence (SEQ ID NO:25)
MKWANQTAVTEYVLMGLHEHCNLEVVLFVFCLGIYSVNVLGNALLIGLNVLHPRLHNPMYFLLSNLS LMDICGTSSFVPLMLDNFLETQRTISFPGCALQMYLTLALGSTECLLLAVMAYDRYVAICQPLRYPE LMSGQTCMQMAALSWGTGFANSLLQSILVWHLPFCGHVINYFYEILAVLKLACGDISLNALALMVAT AVLTLAPLLLICLSYLFILSAILRVPSAAGRCKAFSTCSAHRTVVVVFYGTISFMYFKPKAKDPNVD KTVALFYGVVTPSLNPIIYSLRNAEVKAAVLTLLRGGLLSRKASHCYCCPLPLSAGIG

5   **GPCR9b**

In an alternative embodiment, a GPCR9 variant is the novel GPCR9b (alternatively

referred to herein as 21629632_EXT, Spliced AL133410), which includes the 1069 nucleotide

sequence (SEQ ID NO:26) shown in Table 9C. The GPCR9b ORF begins with a Kozak

consensus ATG initiation codon at nucleotides 31-33 and ends with a TGA codon at

10   nucleotides 1009-1011, which are in bold letters in Table 9C. Putative untranslated regions

upstream from the initiation codon and downstream from the termination codon are underlined

in Table 9A, and the start and stop codons are in bold letters.

Table 9C. GPCR9b Nucleotide Sequence (SEQ ID NO:26)
TGGTCCAGATGGAAAAGGAGAGAGAATGAGATGAAGTGGGCAAACCAGACAGCTGTGACGGAATACG TCCTGATGGGGCTACACGAGCACTGTAACCTGGAGGTGGTCCTGTTTGTGTTCTGCCTGGGCATCTA CTCCGTGAATGTGTTGGGGAACGCCCTCCTCATAGGGCTGAACGTGCTGCACCCTCGCCTGCACAAC CCCATGTACTTCCTTCTCAGCAACCTCTCCCTCATGGACATCTGCGGCACCTCCTCCTTTGTGCCTC TCATGCTAGACAATTTCCTGGAAACCCAGAGGACCATTTCCTTCCCTGGCTGTGCCCTGCAGATGTA CCTGACCCTGGCGCTGGGATCAACGGAGTGCCTGCTGCTGGCTGTGATGGCATATGACCGTTATGTG GCTATCTGCCAGCCGCTTAGGTACCCAGAGCTCATGAGTGGGCAGACCTGCATGCAGATGGCAGCGC TGAGCTGGGGGACAGGCTTTGCCAACTCACTGCTACAGTCCATCCTTGTCTGGCACCTCCCCTTCTG TGGCCACGTCATCAACTACTTCTATGAGATCTTGGCAGTGCTAAAACTGGCCTGTGGGGACATCTCC CTCAATGCGCTGGCATTAATGGTGGCCACAGCCGTCCTGACACTGGCCCCCCTCTTGCTCATCTGCC TGTCTTACCTTTTCATCCTGTCTGCCATCCTTAGGGTACCCTCTGCTGCAGGCCGGTGCAAAGCCTT CTCCACCTGCTCAGCCCACCGCACAGTGGTGGTGGTTTTTTATGGGACAATCTCCTTCATGTACTTC AAACCCAAGGCCAAGGATCCCAACGTGGATAAGACTGTCGCATTGTTCTACGGGGTTGTGACGCCCT CGCTGAACCCCATCATTTACAGCCTGAGGAATGCAGAGGTGAAAGCTGCCGTCCTAACTCTGCTGAG AGGAGGTTTGCTCTCCAGGAAAGCATCCCACTGCTACTGCTGCCCTCTGCCCCTGTCAGCTGGCATA GGCTAGGTTGTGCTGTGGTCATGACCTCAAACCTTGAGAGGCTTAAAGCCATTAAGGTTTGTTT

15   The GPCR9 protein encoded by SEQ ID NO:26 is identical to SEQ ID NO:25.

GPCR9 Clones

Unless specifically addressed as GPCR9a or GPCR9b, any reference to GPCR9 is

assumed to encompass all variants. The GPCR9 nucleic acid sequences differ where GPCR9a

20   extends further in both the 4' and 3' untranslated regions..

In a search of sequence databases, it was found, for example, that the GPCR9b nucleic acid sequence has 856 of 1069 (80%) identical to a *Mouse* Olfactory receptor mRNA (GENBANK-ID: MMU133430). The full GPCR9 amino acid sequence was found to have 231 of 310 amino acid residues (74%) identical to, and 249 of 310 residues (80%) similar to, the 315 amino acid residue protein from *Mouse* (ptnr: SPTREMBL-ACC: Q9QZ17). Additional BLAST results are shown in Table 9E.

Table 9E. BLAST results for GPCR9					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|11464983|ref| NP_062358.1|; gi|5869927|emb|CAB55598.1 (AJ133430); gi|8919698|emb|CAB96153.1 | (AJ251155)	olfactory receptor 70 [Mus musculus]	315	206/315 (65%)	234/315 (73%)	5e-99
gi|11276079|ref| NP_062348.1|; gi|5869920|emb|CAB55594.1 | (AJ133426); gi|8919693|emb|CAB96148.1 | (AJ251154)	olfactory receptor 37c [Mus musculus]	318	170/310 (54%)	207/310 (65%)	2e-77
gi|11276077|ref|NP_06; 2347.1|; gi|5869918|emb|CAB55593.1 | (AJ133425); gi|8919694|emb|CAB96149.1 | (AJ251154)	olfactory receptor 37b [Mus musculus]	318	169/311 (54%)	209/311 (66%)	2e-77
gi|11276075|ref| NP_062346.1|; gi|5869916|emb|CAB55592.1 | (AJ133424); gi|8919692|emb|CAB96147.1 | (AJ251154)	olfactory receptor 37a [Mus musculus]	319	167/312 (53%)	211/312 (67%)	5e-76
gi|11464981|ref| NP_062349.1|; gi|5869923|emb|CAB55596.1 | (AJ133428); gi|8919695|emb|CAB96151.1 | (AJ251154)	olfactory receptor 37e [Mus musculus]	319	165/305 (54%)	203/305 (66%)	3e-74

A multiple sequence alignment is given in Table 9F, with the GPCR9 protein of the invention being shown on line 1, in a ClustalW analysis comparing GPCR9 with related protein sequences, shown in Table 9E.

Table 9F. Information for the ClustalW proteins:

1. GPCR9; SEQ ID NO:25
2. gi|11464983|ref|NP_062358.1| olfactory receptor 70 [Mus musculus]; SEQ ID NO:70
3. gi|11276079|ref|NP_062348.1| olfactory receptor 37c [Mus musculus]; SEQ ID NO:71
4. gi|11276077|ref|NP_062347.1| olfactory receptor 37b [Mus musculus]; SEQ ID NO:72
5. gi|11276075|ref|NP_062346.1| olfactory receptor 37a [Mus musculus]; SEQ ID NO:73
6. gi|11464981|ref|NP_062349.1| olfactory receptor 37e [Mus musculus]; SEQ ID NO:74

```
                        10        20        30        40        50
                    ....|....|....|....|....|....|....|....|....|....|.
    GPCR9           MKWANQTA-VTEYVLMGLHEHCNLEVVLFVFCLGIYSVNVLGNALLIGLN    49
    NP_062358       MAGTNHTE-VIEYVLLGLQDHHGLEIALFVLCLGIYCMTLLGNSFLVGLI    49
 5  NP_062348       MDVSNQTT-VTEFVLLGLSAHPKLEKTFFVLILSMYLVILLGNGVLILVS    49
    NP_062347       MEGANQST-VAEFVLLGLSDHPKLEKTFFVLILLMYLVILLGNGVLILVS    49
    NP_062346       MDRSNETAPLSGFILLGLSAHPKLEKTFFVLILMMYLVILLGNGVLILVS    50
    NP_062349       MERSNKITPVSSFILLGLSAHPKLEKTFFVLILLMYLVILLGNVVLILVS    50

                        60        70        80        90       100
                    ....|....|....|....|....|....|....|....|....|....|.
10  GPCR9           VLHPRLHNPMYFLLSNLSLMDICGTSSFVPLMLDNFLETQRTISFPGCAL    99
    NP_062358       VLDTHLHSPMYFFISNLSLIDICGTSSFTPMMLLNFLDVQRTISFPSCAL    99
    NP_062348       ILDSHLHTPMYFFLGNLSFLDICYTTSSVPLVLDGFLTPRKTISFSGCAV    99
15  NP_062347       ILDSHLHTPMYFFLGDLSFLDICYTTSSIPLVLDGFLTPRKTISFSGCAV    99
    NP_062346       ILDSHLHTPMYFFLGNLSFLDICYTTSSVPLILDSFLTPRKTISFSGCAV   100
    NP_062349       ILDSHLHTPMYFFLGNLSFLDICYTTSSVPLILDSFLTPRKTISFSGCAV   100

                       110       120       130       140       150
                    ....|....|....|....|....|....|....|....|....|....|.
20  PCR9            QMYLTLALGSTECLLLAVMAYDRYVAICQPLRYPELMSGQTCMQMAALSW   149
    NP_062358       QMYLTLALGTTECLLLAVMAYDRYVAICQPLRYPELVNGRYASRWQDK-L   148
    NP_062348       QMFLSFAMGATECVLLGMMAFDRYVAICNPLRYPVVMNKAAYVPMAVSSW   149
25  NP_062347       QMFLSFAMGATECVLLGMMAFDRYVAICNPLRYPVVMNKSAYVPMAASSW   149
    NP_062346       QMFLSFAMGATECVLLSMMAFDRYVAICNPLRYPVVMNKAAYVPMAASSW   150
    NP_062349       QMFLSFAMGATECVLLGMMAFDRYVAICNPLRYPVVMSKAAYVPMAAGSW   150

                       160       170       180       190       200
                    ....|....|....|....|....|....|....|....|....|....|.
30  GPCR9           GTGFANSILQSILVWHLPFCGHVIN--YFYEILAVLKLACGDISLNALAL   197
    NP_062358       GTGFANSLLHSILVWHLPFCGHYIINHFFCEILAVLKLACGDISLNALIL   198
    NP_062348       VAGGANSLVQISLAVQLPFCGDNVINHFICEILAVLKLACADISINVISM   199
35  NP_062347       VAGGANSLVQISLAVQLPFCGDNVINHFTCEILAVLKLACADISINVISM   199
    NP_062346       AGGITNSVVQTSLAMRLPFCGDNVINHFTCEILAVLKLACADISINVISM   200
    NP_062349       VSGSITATVQISLAMTLPFCGDNVINHFTCEILAVLKLACADISINVISM   200

                       210       220       230       240       250
                    ....|....|....|....|....|....|....|....|....|....|.
40  GPCR9           MVATAVLTLAPLLLICLSYLFILSAILRVPSAAGRCKAFSTCSAHRTVVV   247
    NP_062358       TVATAVLTMTPLLLICLSYIFILAAILRVPSAAGRSKAFSTCSAHLTVVV   248
    NP_062348       GVANVIFLGVPVLFIFVSYIFILSTILRIPSAEGRKKAFSTCSAHLTVVI   249
45  NP_062347       GVANVIFLGVPVLFIFVSYIFILSTILRIPSAEGRKKAFSTCSAHLTVVI   249
    NP_062346       VVANMIFLAVPVLFIFVSYVFILVTILRIPSAEGRKKAFSTCSAHLTVVL   250
    NP_062349       AVANAMFLGVPVLFIFVSYIFILSTILRIPSAEGRKKAFSTCSAHLTVVI   250

                       260       270       280       290       300
                    ....|....|....|....|....|....|....|....|....|....|.
50  GPCR9           VFYGTISFMYFKPKAKDPN-------VDKTVALFYGVVTPSLNPIIYSLR   290
    NP_062358       IFYGTITFMYLKPKDQDPS-------VGKITLLYATVAPSLNAFIYSLR   291
    NP_062348       IFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGLLTPMLNPIIYSLR   299
55  NP_062347       VFYGTILFMYGKPKSKDPLGADKQDVSDKLISLFYGVLTPMLNPIIYSLR   299
    NP_062346       VFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGVVTPMLNPIIYSLR   300
    NP_062349       VFYGTILFMYGKPKSKDPLGADKQDLADKLISLFYGVVTPMLNPIIYSLR   300

                       310       320       330
                    ....|....|....|....|....|....|....|.
60
```

```
GPCR9      NAEVKAAVLTLLRGGLLSRKASHCYCCPLPLSAGIG  326
NP_062358  NSEVKAAVTALLWGGLLTRKMSHF------------  315
NP_062348  NKDVKAAVRNLASHRCLTF-----------------  318
NP_062347  NKDVKAAVRNLVGQKCLIQ-----------------  318
NP_062346  NKDVRAAVRNLVGQKHLTE-----------------  319
NP_062349  NKDVKAAVTNLVGQKHFKW-----------------  319
```

DOMAIN results for GPCR9 were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST. This BLAST samples domains found in the Smart and Pfam collections. The results are listed in Table 9G with the statistics and domain description. The 7tm_1, a seven transmembrane receptor (rhodopsin family), was shown to have significant homology to GPCR9. An alignment of GPCR9 residues 41-287 (SEQ ID NO:26) with 7tm_1 residues 1-254 (SEQ ID NO:34) are shown in Table 9G.

Table 9G. DOMAIN results for GPCR9

PSSMs producing significant alignments:	Score (bits)	E value
gnl|Pfam|pfam00001 7tm_1, 7 transmembrane receptor (rhodopsin family)	88.2	6e-19

```
                  10        20        30        40        50
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        GNALLIGLNVLHPRLHNPMYFLLSNLSLMDICGTSSFVPLMLDNFLETQR
Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADLLFLLTLPPWALYYLVGGDW

                  60        70        80        90       100
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        TISFPGCALQMYLTIALGSTECLLLAVMAYDRYVAICQPLRYPELMSGQT
Pfam|pfam00001 VFGDALCKLVGALFVVNGYASILLLTATSIDRYLAIVHPLRYRRTRTPRR

                 110       120       130       140       150
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        CMQMAALSWGTGFANSLLQSLLVW-------HL----PFCGHVINYFYEI
Pfam|pfam00001 AKVLILLVWVLALLLSLPPLLFSWLRTVEEGNT----TVCLIDFPEESVK

                 160       170       180       190       200
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        LAVIKLACGDISLNALALMVATAVLTLAPLLLICLSYLFILS--------
Pfam|pfam00001 RSYVLLSTLVG-----FVLPLLVLVCYTRILRTLRKRARSQ--------

                 210       220       230       240       250
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        --------------------------------------------------
Pfam|pfam00001 --------------------------------------------------

                 260       270       280       290       300
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        ---------------------------------------AILRVPS
Pfam|pfam00001 ---------------------------------------RSLKRRS

                 310       320       330       340       350
             ....|....|....|....|....|....|....|....|....|....|
GPCR9        AAGRCKAFSTCSAHRTVVVVFYGTISFMYFKPKAKD-----PNVDKTVAL
Pfam|pfam00001 SSERKAAKMLLVVVVVFVLCWLPYHIVLLLDSLCLLSIWRVLPTALLITL
```

```
                              360
                     ....|....|....
GPCR9                FYGVVTPSLNPIIY
Pfam|pfam00001       WLAYVNSCLNPIIY
```

The GPCR disclosed in this invention is expressed in at least the following tissues: Prostate, ovary.

The nucleic acids and proteins of GPCR2 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further above.

The novel nucleic acid encoding the GPCR-like protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR2 epitope is from about amino acids 5 to 20. In other specific embodiments, GPCR2 epitopes are from about amino acids 125 to 145, from about amino acids 230 to 240 and from about amino acids 255 to 275.

GPCR10

The disclosed novel GPCR10 nucleic acid (SEQ ID NO:27) of 1147 nucleotides (also referred to as 1823044_EXT) encoding a novel olfactory receptor-like protein is shown in Table 10A. The GPCR disclosed in this invention maps to chromosome 1. An ORF begins with an ATG initiation codon at nucleotides 17-19 and ends with a TAG codon at nucleotides 1061-1063. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 10A, and the start and stop codons are in bold letters.

Table 10A. GPCR10 Nucleotide Sequence (SEQ ID NO:27)
GCCTAGGTGAAACCTCATGGACAACATCACCTGGATGGCCAGCCACACTGGATGGTCGGAT
TTCATCCTGATGGGACTCTTCAGACAATCCAAACATCCAATGGCCAATATCACCTGGATGG
CCAACCACACTGGATGGTCGGATTTCATCCTGTTGGGACTCTTCAGACAATCCAAACATCC
AGCACTACTTTGTGTGGTCATTTTTTGTGGTTTTCCTGATGGCGTTGTCTGGAAATGCTGTC
CTGATCCTTCTGATACACTGTGACGCCCACCTCCACACCCCCATGTACTTTTTCATCAGTC
AATTGTCTCTCATGGACATGGCGTACATTTCTGTCACTGTGCCCAAGATGCTCCTGGACCA
GGTCATGGGTGTGAATAAGATCTCAGCCCCTGAGTGTGGGATGCAGATGTTCTTCTACGTG

```
ACACTAGCAGGTTCAGAATTTTTCCTTCTAGCCACCATGGCCTATGACCGCTACGTGGCCA
TCTGCCATCCTCTCCGTTACCCTGTCCTCATGAACCATAGGGTGTGTCTCTTCCTGTCATC
AGGCTGCTGGTTCCTGGGCTCAGTGGATGGCTTCACATTCACTCCCATCACCATGACCTTC
CCCTTCCGTGGATCCCGGGAGATTCATCATTTCTTCTGTGAAGTTCCTGCTGTATTGAATC
TCTCCTGCTCAGACACCTCACTCTATGAGATTTTCATGTACTTGTGCTGTGTCCTCATGCT
CCTCATCCCTGTGGTGATCATTTCAAGCTCCTATTTACTCATCCTCCTCACCATCCACGGG
ATGAACTCAGCAGAGGGCCGGAAAAAGGCCTTTGCCACCTGCTCCTCCACCTGACTGTGG
TCATCCTCTTCTATGGGGCTGCCATCTACACCTACATGCTCCCCAGCTCCTACCACACCCC
TGAGAAGGACATGATGGTATCTGTCTTCTATACCATCCTCACTCCAGTGGTGAACCCTTTA
ATCTATAGTCTTAGGAATAAGGATGTCATGGGGGCTCTGAAGAAAATGTTAACAGTGGAAC
CTGCCTTTCAAAAAGCTATGGAGTAGACCATTTTGAGAGTAATTTACTTTTCCTTCTCTCT
GCACTTCACATATGAGAATGTTATACCAGTGTTATTTCCCAGACTCCAA
```

The GPCR10 protein (SEQ ID NO:28) encoded by SEQ ID NO:27 has 348 amino acid residues and is presented using the one-letter code in Table 10B. The predicted molecular weight of GPCR10 protein is approximately 39411.93 Daltons. The Psort profile for GPCR10

5　　predicts that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6. In alternative embodiments, GPCR10 is located in the golgi body with a certainty of 0.4, the endoplasmic reticulum (membrane) with a certainty of 0.3 or microbodies (peroxisomes) with a certainty of 0.3. The Signal P predicts a likely cleavage site between positions 19 and 20, i.e., at the dash in the sequence ILM-GL. The protein predicted

10　　here is similar to the "Olfactory Receptor-Like Protein Family", some members of which end up localized at the cell surface where they exhibit activity. Therefore, it is likely that this novel olfactory receptor-like proteins is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

Table 10B. Encoded GPCR10 protein sequence (SEQ ID NO:28)
MDNITWMASHTGWSDFILMGLFRQSKHPMANITWMANHTGWSDFILLGLFRQSKHPALLC VVIFVVFLMALSGNAVLILLIHCDAHLHTPMYFFISQLSLMDMAYISVTVPKMLLDQVMG VNKISAPECGMQMFFYVTLAGSEFFLLATMAYDRYVAICHPLRYPVLMNHRVCLFLSSGC WFLGSVDGFTFTPITMTFPFRGSREIHHFFCEVPAVLNLSCSDTSLYEIFMYLCCVLMLL IPVVIISSSYLLILLTIHGMNSAEGRKKAFATCSSHLTVVILFYGAAIYTYMLPSSYHTP EKDMMVSVFYTILTPVVNPLIYSLRNKDVMGALKKMLTVEPAFQKAME

15　　In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention has 434 of 488 bases (88%) identical to a *gibbon* olfactory receptor mRNA (GENBANK-ID: AF179779). The full amino acid sequence of the protein of the invention was found to have 147 of 223 amino acid residues (65%) identical to, and 177 of 223 residues (79%) similar to the 223 amino acid residue protein from *mouse*

(ptnr:SPTREMBL-ACC: Q62342). GPCR10 also has homology to the proteins shown in the
BLASTP data in Table 10C.

Table 10C. BLAST results for GPCR10					
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|14423768|sp|O43869	O2T1_HUMAN OLFACTORY RECEPTOR 2T1 (OLFACTORY RECEPTOR 1-25) (OR1-25)	311	165/300 (55%)	202/300 (67%)	2e-83
gi|12007424|gb|AAG451 97.1| (AF321234)	T3 olfactory receptor [Mus musculus]	315	139/303 (45%)	186/303 (60%)	3e-68
gi|12007423|gb|AAG451 96.1| (AF321234)	T2 olfactory receptor [Mus musculus]	316	138/301 (45%)	184/301 (60%)	8e-68
gi|12856092|dbj|BAB30 564.1| (AK017036)	putative [Mus musculus]	316	134/301 (44%)	186/301 (61%)	2e-66
gi|12855358|dbj|BAB30 304.1| (AK016560)	putative [Mus musculus]	316	134/301 (44%)	186/301 (61%)	2e-66

A multiple sequence alignment is given in Table 10D, with the protein of the invention
1823044_EXT being shown on lines 1 in Table 10D in a ClustalW analysis comparing the
protein of the invention with related protein sequences. This BLASTP data is displayed
graphically in the Clustal W in Table 10D.

Table 10D. ClustalW Analysis of GPCR10

1) GPCR10; SEQ ID NO:28
2) >gi\|14423768\|sp\|O43869\|O2T1_ Human Olfactory Receptor 2T1 (OR1-25) ; SEQ ID NO:75
3) >gi\|12007424\|gb\|AAG45197.1\| (AF321234) T3 olfactory receptor [Mus musculus]; SEQ ID NO:76
4) >gi\|12007423\|gb\|AAG45196.1\| (AF321234) T2 olfactory receptor [Mus musculus]; SEQ ID NO:77
5) >gi\|12856092\|dbj\|BAB30564.1\| (AK017036) putative [Mus musculus]; SEQ ID NO:78
6) >gi\|12855358\|dbj\|BAB30304.1\| (AK016560) putative [Mus musculus]; SEQ ID NO:79

```
                        10        20        30        40        50
                 ....|....|....|....|....|....|....|....|....|....|
GPCR10           MDNITWMASHTGWSDFILMGLFRQSKHPMANITWMANHTGWSDFILLGLF
gi|14423768|     ------------------------------------MEEYNTSSTDFTFMGLF
gi|12007424|     -----------------------------------MEVCNSTLRSGFILMGIL
gi|12007423|     -----------------------------------MEPWNSTLGTDFNLVGIL
gi|12856092|     -----------------------------------MEPWNSTLESGFILVGIL
gi|12855358|     -----------------------------------MEPWNSTLESGFILVGIL

                        60        70        80        90        100
                 ....|....|....|....|....|....|....|....|....|....|
GPCR10           RQSKHPALLCVVLFVVFLMALSGNAVLILLIHCDAHLHTPMYFFISQLSL
gi|14423768|     NRKETSGLIFAILSILFFTALMANGVMIFLIQTDLRLHTPMYFLLSHLSL
gi|12007424|     DDNDFPELLCATITALYLLALTSNGLLLLVITMDTRLHVPMYLLLWQLSL
gi|12007423|     DDSGSPELLCATFTALYMLALISNGLLILVITMDARLHVPMYFLLGQLSL
gi|12856092|     DGSGSPELLCATVTTLYMLALISNGLLLLVITVDARLHVPMYLLLRQLSL
```

64

gi|12855358| DGSGSPELLCATVTTLYMLALISNGLLLLVITVDARLHVPMYLLLRQLSL

```
                    110        120        130        140        150
                ....|....|....|....|....|....|....|....|....|....|
 5   GPCR10     MDMAYISVTVPKMLLDQVMGVNKISAPECGMQMFFYVTLAGSEFFLLATM
     gi|14423768| IDMMYISTTVPKMLVNYLLDQRTISFVGCTAQHFLYLTLVGAEFFLLGLM
     gi|12007424| MDLLLTSVITPKAILDYLLKDNTISFGGCALQMFLALTLGTAEDLLLSFM
     gi|12007423| MDLLFTSVVTPKAVIDFLLRDNTISFEGCSLQMFLALTLGGAEDLLLAFM
     gi|12856092| IDLLFTSVVTPNTVVDFLLRDNTISFEGCALQLFSAMTLGGAEDLLLAFM
10   gi|12855358| IDLLFTSVVTPNTVVDFLLRDNTISFEGCALQLFSAMTLGGAEELLLAFM

                    160        170        180        190        200
                ....|....|....|....|....|....|....|....|....|....|
     GPCR10     AYDRYVAICHPLRYPVLMNHRVCLFLSSGCWFLGSVDGFTFTPITMTFPF
15   gi|14423768| AYDRYVAICNPLRYPVLMSRRVCWMIIAGSWFGGSLDGFLLTPITMSFPF
     gi|12007424| AYDRYVAICHPLNYTILMSQKVCCLMIATSWSLASLSALGYSMYTMQYPF
     gi|12007423| AYDRYVAICHPLNYMIFMRPSICWLMVATSWVLASLMALGYTTYTMQYSY
     gi|12856092| AYDRYVAICHPLNYMIFMSPKACRLMVAISWILASLSALGHTVYTMHFPF
     gi|12855358| AYDRYVAICHPLNYMIFMSPKACRLMVAISWILASLSALGHTVYTMHFPF
20
                    210        220        230        240        250
                ....|....|....|....|....|....|....|....|....|....|
     GPCR10     RGSREIHHFFCEVPAVLNLSCSDTSLYETFMYLCCVLMLLIPVVIISSSY
     gi|14423768| CNSREINHFFCEAPAVLKLACADTALYETVMYVCCVLMLLIPFSVVLASY
25   gi|12007424| CKSRQIRHLFCEIPPLLKLACADTSYELMVYLMGVTLLFPALAAILASY
     gi|12007423| CKSRKIRHLLCEIPPLLKLACADTSKYELMVYVMGVTFLTPPLAAILASY
     gi|12856092| CMSQEIRHLLCEVPPLLKLACADTSQYELMVYVTGVIFLLLPLSAIETSY
     gi|12855358| CMSQEIRHLLCEVPPLLKLACADTSQYELMVYVTGVIFLLLPLSAILTSY
30
                    260        270        280        290        300
                ....|....|....|....|....|....|....|....|....|....|
     GPCR10     LLIILTIHGMNSAEGRKKAFATCSSHLTVVILFYGAAIYTYMLPSSYHTP
     gi|14423768| ARIITTVQCMSSVEGRKKAFATCSSHMTVVSLFYGAAMYIYMLPHSYHKP
     gi|12007424| SLILFTVLHMPSNEGRRKALVTCSSHLTVVGMWYGGAIVMYVLPSSFHSP
35   gi|12007423| SLILFTVLHMPSNEGRKKALVTCSSHLTVVGMFYGAATFMYVLPNSFHSP
     gi|12856092| SLILFTVLHMPSNEGRKKALVTCSSHLTVVGMFYGGATFMYVLPSSFHSP
     gi|12855358| SLILFTVLHMPSNEGRKKALVTCSSHLTVVGMFYGGATFMYVLPSSFHSP

                    310        320        330        340
                ....|....|....|....|....|....|....|....|....|...
40   GPCR10     EKDMMVSVFYTILTPVVNPLIYSLRNKDVMGALKKMLTVEPAFQKAME
     gi|14423768| AQDKVLSVFYTILTPMLNPLIYSLRNKDVTGALKRALGRFKGPQ----
     gi|12007424| KQDNISSVFYTIFTPALNPLIYSLRNKEVTGALRRVLGKRLSVQSTF-
     gi|12007423| RQDNISVFYTIVTPALNPLIYSLRNKEVTGALIRVLGRYTVPAHPTL
45   gi|12856092| KQDNIISVFYTIVTPALNPLIYSLRNKEVIGAVRRVLGRHLLPAHATV
     gi|12855358| KQDNIISVFYTIVTPALNPLIYSLRNKEVIGAVRRVLGRHLLPAHATV
```

Table 10E lists the domain description from DOMAIN analysis results against
GPCR10. This indicates that the GPCR10 sequence has properties similar to those of other
50   proteins known to contain this domain as well as to the 254 amino acid 7tm domain (SEQ ID
NO:34) itself.

Table 10E Domain Analysis of GPCR10

```
PSSMs producing significant alignments:                           Score  E
                                                                  (bits) value
gnl|Pfam|pfam00001  7tm_1, 7 transmembrane receptor (rhodopsin family) 90.9  1e-19


                          10        20        30        40        50
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        GNAVLILLIHCDAHLHTPMYFFISQLSLMDMAYISVTVPKMLLDQVMG--
     Pfam|pfam00001 GNLLVILVILRTKKLRTPTNIFLLNLAVADILFLLTLPPWALYYLVGG--


                          60        70        80        90        100
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        ----VNKISAPECGMQMFFYVTLAGSEFFLLATMAYDRYVAICHPLRYPV
     Pfam|pfam00001 ----DWVFGDALCKLVGALFVVNGYASILLLTAISIDRYLAIVHPLRYRR


                          110       120       130       140        150
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        LMNHRVCLFLSSGCWFLGSWDGFTFTPIIMTFPFRGSRE----IHHFFCE
     Pfam|pfam00001 IRTPRRAKVLILLVWVLALLLSLPPLLFSWLRTVEEGNT----TVCLIDF


                          160       170       180       190        200
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        VPAVLNLSCSDTSLYEIFMYLCCVLMLLIPVVIISSSYLLILL-------
     Pfam|pfam00001 PEESVKRSYVLLSTLVGFVLPLLVILVCYTRILRTLRKRARSQ-------


                          210       220       230       240        250
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        -------------------------------------------------
     Pfam|pfam00001 -------------------------------------------------


                          260       270       280       290        300
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        ---------------------------------------------TIHGMN
     Pfam|pfam00001 ---------------------------------------------RSLKRR


                          310       320       330       340        350
                   ....|....|....|....|....|....|....|....|....|....|
     GPCR10        SAEGRKKAFATCSSHLTVVLLFYGAAIYTYMLP----SSY-HTPEKDMMV
     Pfam|pfam00001 SSSERKAAKMLLVVVVVFVFLCWLPYHIVLLLDSLCLLSIW-RVLPTALLI


                          360
                   ....|....|....|.
     GPCR10        SVFYTILTPVVNPLIY
     Pfam|pfam00001 TLWLAYVNSCLNPIIY
```

The olfactory receptor disclosed in this invention is expressed in at least the following tissues: lymph node, ovary.

The nucleic acids and proteins of GPCR10 are useful in potential therapeutic applications implicated in various GPCR-related pathological disorders and/or OR-related pathological disorders, described further herein. The novel GPCR10 nucleic acid, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or

diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-GPCRX Antibodies" section below. The disclosed GPCR10 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated GPCR10
5   epitope is from about amino acids 10 to 18. In additional embodiments, GPCR10 epitopes are from about amino acids 20 to 30, from about amino acids 42 to 50, from about amino acids 190 to 210, from about amino acids 260 to 270, from about amino acids 280 to 308 and from about amino acids 325 to 340.

A summary of the GPCRX nucleic acids and proteins of the invention is provided in
10   Table 11.

TABLE 11: Summary Of Nucleic Acids And Proteins Of The Invention

Name	Tables	Clone; Description of Homolog	Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO
GPCR1	1A, 1B	GPCR1a: CG54326_02	1	2
	1C, 1D	GPCR1b: AP001804_A	3	4
GPCR2	2A, 2B	GPCR2a: CG54335_02	5	6
	2C, 2D	GPCR2b: AP001804_B	7	8
GPCR3	3A, 3B	GPCR3: AP001804_C	9	10
GPCR4	4A, 4B	GPCR4: AP001804_D	11	12
GPCR5	5A, 5B,	GPCR5a: CG56040_01	13	14
	5C, 5D	GPCR5b: AP001804_E	15	16
GPCR6	6A, 6B	GPCR6a: CG56025-01	17	18
	6C	GPCR6b: AP001804_B	19	
GPCR7	7A, 7B	GPCR7: 20722608_EXT	20	21
GPCR8	8A, 8B	GPCR8: CG-SC931712	22	23
GPCR9	9A, 9B	GPCR9a: 21629632.0.20	24	25
	9C	GPCR9b: 21629632_E	26	
GPCR10	10A, 10B	GPCR10: 1823044_EXT	27	28

GPCRX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode
15   GPCRX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify GPCRX-encoding nucleic acids (*e.g.*, GPCRX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of GPCRX nucleic acid molecules. As used herein, the term

"nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic

DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using

nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid

molecule may be single-stranded or double-stranded, but preferably is comprised double-

5   stranded DNA.

An GPCRX nucleic acid can encode a mature GPCRX polypeptide. As used herein, a

"mature" form of a polypeptide or protein disclosed in the present invention is the product of a

naturally occurring polypeptide or precursor form or proprotein. The naturally occurring

polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length

10   gene product, encoded by the corresponding gene. Alternatively, it may be defined as the

polypeptide, precursor or proprotein encoded by an ORF described herein. The product

"mature" form arises, again by way of nonlimiting example, as a result of one or more

naturally occurring processing steps as they may take place within the cell, or host cell, in

which the gene product arises. Examples of such processing steps leading to a "mature" form

15   of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded

by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader

sequence. Thus a mature form arising from a precursor polypeptide or protein that has

residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through

N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising

20   from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal

sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to

residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may

arise from a step of post-translational modification other than a proteolytic cleavage event.

Such additional processes include, by way of non-limiting example, glycosylation,

25   myristoylation or phosphorylation. In general, a mature polypeptide or protein may result

from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable

length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as

approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the

30   detection of identical, similar, or complementary nucleic acid sequences. Longer length

probes are generally obtained from a natural or recombinant source, are highly specific, and

much slower to hybridize than shorter-length oligomer probes. Probes may be single- or

double-stranded and designed to have specificity in PCR, membrane-based hybridization

technologies, or ELISA-like technologies.

68

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated GPCRX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 as a hybridization probe, GPCRX molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, et al., (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to GPCRX nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or

100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the

invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length

would further comprise at least 6 contiguous nucleotides of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13,

15, 17, 19, 20, 22, 24, 26 and 27, or a complement thereof. Oligonucleotides may be

5   chemically synthesized and may also be used as probes.

        In another embodiment, an isolated nucleic acid molecule of the invention comprises a

nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:

1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or a portion of this nucleotide sequence

(e.g., a fragment that can be used as a probe or primer or a fragment encoding a biologically-

10   active portion of an GPCRX polypeptide). A nucleic acid molecule that is complementary to

the nucleotide sequence shown in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26

and 27 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID

NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 that it can hydrogen bond with

little or no mismatches to the nucleotide sequence shown SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13,

15   15, 17, 19, 20, 22, 24, 26 and 27, thereby forming a stable duplex.

        As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base

pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means

the physical or chemical interaction between two polypeptides or compounds or associated

polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van

20   der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct

or indirect. Indirect interactions may be through or due to the effects of another polypeptide or

compound. Direct binding refers to interactions that do not take place through, or due to, the

effect of another polypeptide or compound, but instead are without other substantial chemical

intermediates.

25   Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic

acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific

hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of

amino acids, respectively, and are at most some portion less than a full length sequence.

Fragments may be derived from any contiguous portion of a nucleic acid or amino acid

30   sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed

from the native compounds either directly or by modification or partial substitution. Analogs

are nucleic acid sequences or amino acid sequences that have a structure similar to, but not

identical to, the native compound but differs from it in respect to certain components or side

chains. Analogs may be synthetic or from a different evolutionary origin and may have a

similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of GPCRX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an GPCRX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human GPCRX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, as well as a polypeptide possessing GPCRX biological activity. Various biological activities of the GPCRX proteins are described below.

As used herein, "identical" residues correspond to those residues in a comparison between two sequences where the equivalent nucleotide base or amino acid residue in an alignment of two sequences is the same residue. Residues are alternatively described as "similar" or "positive" when the comparisons between two sequences in an alignment show

71

that residues in an equivalent position in a comparison are either the same amino acid or a
conserved amino acid as defined below.

     An GPCRX polypeptide is encoded by the open reading frame ("ORF") of an GPCRX
nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated
5   into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop
codon. An ORF that represents the coding sequence for a full protein begins with an ATG
"start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or
TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with
or without a start codon, a stop codon, or both. For an ORF to be considered as a good
10   candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set,
e.g., a stretch of DNA that would encode a protein of 50 amino acids or more.

     The nucleotide sequences determined from the cloning of the human GPCRX genes
allows for the generation of probes and primers designed for use in identifying and/or cloning
GPCRX homologues in other cell types, *e.g.* from other tissues, as well as GPCRX
15   homologues from other vertebrates. The probe/primer typically comprises substantially
purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide
sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200,
250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7,
9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27; or an anti-sense strand nucleotide sequence of SEQ
20   ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27; or of a naturally occurring
mutant of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27.

     Probes based on the human GPCRX nucleotide sequences can be used to detect
transcripts or genomic sequences encoding the same or homologous proteins. In various
embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group
25   can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such
probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-
express an GPCRX protein, such as by measuring a level of an GPCRX-encoding nucleic acid
in a sample of cells from a subject *e.g.,* detecting GPCRX mRNA levels or determining
whether a genomic GPCRX gene has been mutated or deleted.

30   "A polypeptide having a biologically-active portion of an GPCRX polypeptide" refers
to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a
polypeptide of the invention, including mature forms, as measured in a particular biological
assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-
active portion of GPCRX" can be prepared by isolating a portion SEQ ID NOS: 1, 3, 5, 7, 9,

11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 that encodes a polypeptide having an GPCRX biological activity (the biological activities of the GPCRX proteins are described below), expressing the encoded portion of GPCRX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of GPCRX.

5   **GPCRX Nucleic Acid and Polypeptide Variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 due to degeneracy of the genetic code and thus encode the same GPCRX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19,

10   20, 22, 24, 26 and 27. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28.

In addition to the human GPCRX nucleotide sequences shown in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 it will be appreciated by those skilled in the art

15   that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the GPCRX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the GPCRX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an GPCRX

20   protein, preferably a vertebrate GPCRX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the GPCRX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the GPCRX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the GPCRX polypeptides, are intended to be within the scope of the invention.

25   Moreover, nucleic acid molecules encoding GPCRX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the GPCRX cDNAs of the invention can be isolated based on their homology to the human

30   GPCRX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27. In another embodiment, the nucleic acid is at least 10,

5   25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

10   Homologs (i.e., nucleic acids encoding GPCRX proteins derived from species other than human) or other related sequences (e.g., paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions

15   under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The

20   Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M

25   sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel,

30   et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM

EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 corresponds

5   to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15,

10   17, 19, 20, 22, 24, 26 and 27 or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are

15   well-known within the art. See, e.g., Ausubel, et al. (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22,

20   24, 26 and 27 or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH

25   7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel, et al. (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. Proc Natl Acad Sci USA 78: 6789-6792.

30   **Conservative Mutations**

In addition to naturally-occurring allelic variants of GPCRX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 thereby leading to changes in the amino acid sequences of the encoded

GPCRX proteins, without altering the functional ability of said GPCRX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28. A "non-essential" amino acid residue is a residue that can be altered from the

5   wild-type sequences of the GPCRX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the GPCRX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

10   Another aspect of the invention pertains to nucleic acid molecules encoding GPCRX proteins that contain changes in amino acid residues that are not essential for activity. Such GPCRX proteins differ in amino acid sequence from SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein

15   comprises an amino acid sequence at least about 45% homologous to the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28; more preferably at least about 70% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28; still more

20   preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28; and most preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28.

An isolated nucleic acid molecule encoding an GPCRX protein homologous to the

25   protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

30   Mutations can be introduced into SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side

chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side

5   chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the GPCRX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced

10   randomly along all or part of an GPCRX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for GPCRX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

15   The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be

20   substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

            In one embodiment, a mutant GPCRX protein can be assayed for (*i*) the ability to form protein:protein interactions with other GPCRX proteins, other cell-surface proteins, or

25   biologically-active portions thereof, (*ii*) complex formation between a mutant GPCRX protein and an GPCRX ligand; or (*iii*) the ability of a mutant GPCRX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

            In yet another embodiment, a mutant GPCRX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

30   **Antisense Nucleic Acids**

            Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or

fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that

5   comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire GPCRX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an GPCRX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, or antisense nucleic acids complementary to an GPCRX nucleic acid sequence of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22,

10   24, 26 and 27, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an GPCRX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid

15   molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the GPCRX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the GPCRX protein disclosed herein,

20   antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of GPCRX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of GPCRX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the

25   translation start site of GPCRX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or

30   variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine,

xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine,

5   7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil,

10 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

15   The antisense nucleic acid molecules of the invention are typically administered to a subject or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an GPCRX protein to thereby inhibit expression of the protein (e.g., by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an

20 antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified

25 such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong

30 pol II or pol III promoter are preferred.

          In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15:

6625-6641. The antisense nucleic acid molecule can also comprise a

2'-o-methylribonucleotide (see, *e.g.,* Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a

chimeric RNA-DNA analogue (see, *e.g.,* Inoue, *et al.,* 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

5   Nucleic acid modifications include, by way of non-limiting example, modified bases,

and nucleic acids whose sugar phosphate backbones are modified or derivatized. These

modifications are carried out at least in part to enhance the chemical stability of the modified

nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in

therapeutic applications in a subject.

10   In one embodiment, an antisense nucleic acid of the invention is a ribozyme.

Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of

cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a

complementary region. Thus, ribozymes (*e.g.,* hammerhead ribozymes as described in

Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave GPCRX

15   mRNA transcripts to thereby inhibit translation of GPCRX mRNA. A ribozyme having

specificity for an GPCRX-encoding nucleic acid can be designed based upon the nucleotide

sequence of an GPCRX cDNA disclosed herein (*i.e.,* SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15,

17, 19, 20, 22, 24, 26 and 27). For example, a derivative of a *Tetrahymena* L-19 IVS RNA

can be constructed in which the nucleotide sequence of the active site is complementary to the

20   nucleotide sequence to be cleaved in an GPCRX-encoding mRNA. See, *e.g.,* U.S. Patent

4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* GPCRX mRNA can also

be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA

molecules. See, *e.g.,* Bartel *et al.,* (1993) *Science* 261:1411-1418.

          Alternatively, GPCRX gene expression can be inhibited by targeting nucleotide

25   sequences complementary to the regulatory region of the GPCRX nucleic acid (*e.g.,* the

GPCRX promoter and/or enhancers) to form triple helical structures that prevent transcription

of the GPCRX gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84;

Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

          In various embodiments, the GPCRX nucleic acids can be modified at the base moiety,

30   sugar moiety or phosphate backbone to improve, *e.g.,* the stability, hybridization, or solubility

of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can

be modified to generate peptide nucleic acids. *See, e.g.,* Hyrup, *et al.,* 1996. *Bioorg Med*

Chem 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic

acid mimics (*e.g.,* DNA mimics) in which the deoxyribose phosphate backbone is replaced by

a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*;

5   Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of GPCRX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of GPCRX can also be used, for example, in the analysis of single base pair mutations in a

10 gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*see*, Hyrup, *et al.*, 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*see*, Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.*, 1996. *supra*).

In another embodiment, PNAs of GPCRX can be modified, *e.g.*, to enhance their

15 stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of GPCRX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion

20 while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (*see*, Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.*, 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can

25 be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.*, Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.*,

30 Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.*, Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across

the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86:

6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No.

WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In

addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see,*

5   *e.g.*, Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988.

Pharm. Res. 5: 539-549). To this end, the oligonucleotide may be conjugated to another

molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a

hybridization-triggered cleavage agent, and the like.

GPCRX Polypeptides

10   A polypeptide according to the invention includes a polypeptide including the amino

acid sequence of GPCRX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6,

8, 10, 12, 14, 16, 18, 21, 23, 25 and 28. The invention also includes a mutant or variant

protein any of whose residues may be changed from the corresponding residues shown in SEQ

ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28 while still encoding a protein that

15   maintains its GPCRX activities and physiological functions, or a functional fragment thereof.

In general, an GPCRX variant that preserves GPCRX-like function includes any

variant in which residues at a particular position in the sequence have been substituted by

other amino acids, and further include the possibility of inserting an additional residue or

residues between two residues of the parent protein as well as the possibility of deleting one or

20   more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is

encompassed by the invention. In favorable circumstances, the substitution is a conservative

substitution as defined above.

One aspect of the invention pertains to isolated GPCRX proteins, and biologically-

active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided

25   are polypeptide fragments suitable for use as immunogens to raise anti-GPCRX antibodies. In

one embodiment, native GPCRX proteins can be isolated from cells or tissue sources by an

appropriate purification scheme using standard protein purification techniques. In another

embodiment, GPCRX proteins are produced by recombinant DNA techniques. Alternative to

recombinant expression, an GPCRX protein or polypeptide can be synthesized chemically

30   using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof

is substantially free of cellular material or other contaminating proteins from the cell or tissue

source from which the GPCRX protein is derived, or substantially free from chemical

precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of GPCRX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes

5   preparations of GPCRX proteins having less than about 30% (by dry weight) of non-GPCRX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-GPCRX proteins, still more preferably less than about 10% of non-GPCRX proteins, and most preferably less than about 5% of non-GPCRX proteins. When the GPCRX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably

10   substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the GPCRX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of GPCRX proteins in which the protein is separated from chemical precursors or

15   other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of GPCRX proteins having less than about 30% (by dry weight) of chemical precursors or non-GPCRX chemicals, more preferably less than about 20% chemical precursors or non-GPCRX chemicals, still more preferably less than about 10% chemical precursors or

20   non-GPCRX chemicals, and most preferably less than about 5% chemical precursors or non-GPCRX chemicals.

Biologically-active portions of GPCRX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the GPCRX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14,

25   16, 18, 21, 23, 25 and 28) that include fewer amino acids than the full-length GPCRX proteins, and exhibit at least one activity of an GPCRX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the GPCRX protein. A biologically-active portion of an GPCRX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

30   Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native GPCRX protein.

In an embodiment, the GPCRX protein has an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28. In other embodiments, the GPCRX

protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the GPCRX

5   protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, and retains the functional activity of the GPCRX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28.

Determining Homology Between Two or More Sequences

10   To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a

15   position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity

20   between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, Needleman and Wunsch, 1970. J Mol Biol 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above

25   exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of

30   comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the

number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80

5   percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides GPCRX chimeric or fusion proteins. As used herein, an

10   GPCRX "chimeric protein" or "fusion protein" comprises an GPCRX polypeptide operatively-linked to a non-GPCRX polypeptide. An "GPCRX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an GPCRX protein (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28), whereas a "non-GPCRX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the GPCRX protein, *e.g.*, a protein that is different from the GPCRX protein

15   and that is derived from the same or a different organism. Within an GPCRX fusion protein the GPCRX polypeptide can correspond to all or a portion of an GPCRX protein. In one embodiment, an GPCRX fusion protein comprises at least one biologically-active portion of an GPCRX protein. In another embodiment, an GPCRX fusion protein comprises at least two

20   biologically-active portions of an GPCRX protein. In yet another embodiment, an GPCRX fusion protein comprises at least three biologically-active portions of an GPCRX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the GPCRX polypeptide and the non-GPCRX polypeptide are fused in-frame with one another. The non-GPCRX polypeptide can be fused to the N-terminus or C-terminus of the GPCRX

25   polypeptide.

In one embodiment, the fusion protein is a GST-GPCRX fusion protein in which the GPCRX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant GPCRX polypeptides.

30   In another embodiment, the fusion protein is an GPCRX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of GPCRX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an GPCRX-immunoglobulin fusion protein in which the GPCRX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The GPCRX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject

5   to inhibit an interaction between an GPCRX ligand and an GPCRX protein on the surface of a cell, to thereby suppress GPCRX-mediated signal transduction *in vivo*. The GPCRX-immunoglobulin fusion proteins can be used to affect the bioavailability of an GPCRX cognate ligand. Inhibition of the GPCRX ligand/GPCRX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as

10   modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the GPCRX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-GPCRX antibodies in a subject, to purify GPCRX ligands, and in screening assays to identify molecules that inhibit the interaction of GPCRX with an GPCRX ligand.

An GPCRX chimeric or fusion protein of the invention can be produced by standard

15   recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In

20   another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR

25   BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). An GPCRX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCRX protein.

GPCRX Agonists and Antagonists

30   The invention also pertains to variants of the GPCRX proteins that function as either GPCRX agonists (*i.e.*, mimetics) or as GPCRX antagonists. Variants of the GPCRX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the GPCRX protein). An agonist of the GPCRX protein can retain substantially the same, or a subset of,

the biological activities of the naturally occurring form of the GPCRX protein. An antagonist of the GPCRX protein can inhibit one or more of the activities of the naturally occurring form of the GPCRX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the GPCRX protein. Thus, specific

5   biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the GPCRX proteins.

Variants of the GPCRX proteins that function as either GPCRX agonists (*i.e.*,

10   mimetics) or as GPCRX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the GPCRX proteins for GPCRX protein agonist or antagonist activity. In one embodiment, a variegated library of GPCRX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of GPCRX variants can be produced by, for example,

15   enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential GPCRX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of GPCRX sequences therein. There are a variety of methods which can be used to produce libraries of potential GPCRX variants from a degenerate oligonucleotide sequence. Chemical

20   synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential GPCRX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3;

25 · Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the GPCRX protein coding sequences can be used to generate a variegated population of GPCRX fragments for screening and subsequent

30   selection of variants of an GPCRX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an GPCRX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded

DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the GPCRX

5   proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of GPCRX proteins. The most

10   widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble

15   mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify GPCRX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.,* 1993. *Protein Engineering* 6:327-331.

Anti-GPCRX Antibodies

20   Also included in the invention are antibodies to GPCRX proteins, or fragments of GPCRX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab},

25     $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a

30   reference to all such classes, subclasses and types of human antibody species.

An isolated GPCRX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for

polyclonal and monoclonal antibody preparation. The full-length protein can be used or,

alternatively, the invention provides antigenic peptide fragments of the antigen for use as

immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the

amino acid sequence of the full length protein and encompasses an epitope thereof such that an

5   antibody raised against the peptide forms a specific immune complex with the full length

protein or with any fragment that contains the epitope. Preferably, the antigenic peptide

comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20

amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by

the antigenic peptide are regions of the protein that are located on its surface; commonly these

10   are hydrophilic regions.

        In certain embodiments of the invention, at least one epitope encompassed by the

antigenic peptide is a region of GPCRX-related protein that is located on the surface of the

protein, e.g., a hydrophilic region. A hydrophobicity analysis of the human GPCRX-related

protein sequence will indicate which regions of a GPCRX-related protein are particularly

15   hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody

production. As a means for targeting antibody production, hydropathy plots showing regions

of hydrophilicity and hydrophobicity may be generated by any method well known in the art,

including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without

Fourier transformation. See, e.g., Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78:

20   3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated

herein by reference in its entirety. Antibodies that are specific for one or more domains within

an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided .

herein.

        A protein of the invention, or a derivative, fragment, analog, homolog or ortholog

25   thereof, may be utilized as an immunogen in the generation of antibodies that

immunospecifically bind these protein components.

        Various procedures known within the art may be used for the production of polyclonal

or monoclonal antibodies directed against a protein of the invention, or against derivatives,

fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory

30   Manual, Harlow and Lane, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor,

NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

        For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit,

goat, mouse or other mammal) may be immunized by one or more injections with the native

protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate

immunogenic preparation can contain, for example, the naturally occurring immunogenic

protein, a chemically synthesized polypeptide representing the immunogenic protein, or a

recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated

5   to a second protein known to be immunogenic in the mammal being immunized. Examples of

such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum

albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further

include an adjuvant. Various adjuvants used to increase the immunological response include,

but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum

10   hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions,

peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille

Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents.

Additional examples of adjuvants which can be employed include MPL-TDM adjuvant

(monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

15   The polyclonal antibody molecules directed against the immunogenic protein can be

isolated from the mammal (e.g., from the blood) and further purified by well known

techniques, such as affinity chromatography using protein A or protein G, which provide

primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific

antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be

20   immobilized on a column to purify the immune specific antibody by immunoaffinity

chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson

(The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17,

2000), pp. 25-28).

Monoclonal Antibodies

25   The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as

used herein, refers to a population of antibody molecules that contain only one molecular

species of antibody molecule consisting of a unique light chain gene product and a unique

heavy chain gene product. In particular, the complementarity determining regions (CDRs) of

the monoclonal antibody are identical in all the molecules of the population. MAbs thus

30   contain an antigen binding site capable of immunoreacting with a particular epitope of the

antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those

described by Kohler and Milstein, *Nature*, 256:495 (1975). In a hybridoma method, a mouse,

hamster, or other appropriate host animal, is typically immunized with an immunizing agent to

elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, *J. Immunol.*, 133:3001 (1984); Brodeur et al., MONOCLONAL ANTIBODY PRODUCTION TECHNIQUES AND APPLICATIONS, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, *Anal. Biochem.*, 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

91

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

5   The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as

10   those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors,

15   which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No.

20   4,816,567; Morrison, *Nature* 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

25   **Humanized Antibodies**

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins,

30   immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')$_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., *Nature*, 321:522-525 (1986); Riechmann et al., *Nature*, 332:323-327 (1988); Verhoeyen et al.,

Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which

5   are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally

10   also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire

15   sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL

20   ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

25   In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, *J. Mol. Biol.*, 227:381 (1991); Marks et al., *J. Mol. Biol.*, 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon

30   challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (*Bio/Technology* 10, 779-783 (1992)); Lonberg et al. (*Nature* 368 856-859 (1994)); Morrison (*Nature* 368, 812-13 (1994)); Fishwild

et al,(*Nature Biotechnology* 14, 845-51 (1996)); Neuberger (*Nature Biotechnology* 14, 826

(1996)); and Lonberg and Huszar (*Intern. Rev. Immunol.* 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals

which are modified so as to produce fully human antibodies rather than the animal's

5   endogenous antibodies in response to challenge by an antigen. (See PCT publication

WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in

the nonhuman host have been incapacitated, and active loci encoding human heavy and light

chain immunoglobulins are inserted into the host's genome. The human genes are

incorporated, for example, using yeast artificial chromosomes containing the requisite human

10   DNA segments. An animal which provides all the desired modifications is then obtained as

progeny by crossbreeding intermediate transgenic animals containing fewer than the full

complement of the modifications. The preferred embodiment of such a nonhuman animal is a

mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and

WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins.

15   The antibodies can be obtained directly from the animal after immunization with an

immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively

from immortalized B cells derived from the animal, such as hybridomas producing

monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human

variable regions can be recovered and expressed to obtain the antibodies directly, or can be

20   further modified to obtain analogs of antibodies such as, for example, single chain Fv

molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse,

lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent

No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at

25   least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of

the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain

locus, the deletion being effected by a targeting vector containing a gene encoding a selectable

marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and

germ cells contain the gene encoding the selectable marker.

30   A method for producing an antibody of interest, such as a human antibody, is disclosed

in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a

nucleotide sequence encoding a heavy chain into one mammalian host cell in culture,

introducing an expression vector containing a nucleotide sequence encoding a light chain into

another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

5   binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_ab Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent

10   No. 4,946,778). In addition, methods can be adapted for the construction of F_ab expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_ab fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but

15   not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_ab fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_ab fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

20   Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the

25   recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, *Nature*, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct

30   bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion

preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the

5   immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which

10   are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing

15   large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. F(ab')$_2$ bispecific antibodies). Techniques for generating bispecific antibodies from

20   antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')$_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab'

25   fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

30   Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')$_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to

cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., *J. Immunol.* 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., *Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., *J. Immunol.* 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., *J. Immunol.* 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells

(U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as

5   glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-

10   methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation

15   using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and

20   other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an GPCRX protein is facilitated by generation of hybridomas that bind to the fragment of an GPCRX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an GPCRX protein, or derivatives, fragments, analogs or homologs thereof, are also provided

25   herein.

Anti-GPCRX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an GPCRX protein (e.g., for use in measuring levels of the GPCRX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for GPCRX

30   proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-GPCRX antibody (e.g., monoclonal antibody) can be used to isolate an GPCRX polypeptide by standard techniques, such as affinity chromatography or

immunoprecipitation. An anti-GPCRX antibody can facilitate the purification of natural

GPCRX polypeptide from cells and of recombinantly-produced GPCRX polypeptide

expressed in host cells. Moreover, an anti-GPCRX antibody can be used to detect GPCRX

protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and

5   pattern of expression of the GPCRX protein. Anti-GPCRX antibodies can be used

diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to,

for example, determine the efficacy of a given treatment regimen. Detection can be facilitated

by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of

detectable substances include various enzymes, prosthetic groups, fluorescent materials,

10   luminescent materials, bioluminescent materials, and radioactive materials. Examples of

suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or

acetylcholinesterase; examples of suitable prosthetic group complexes include

streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include

umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine

15   fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes

luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and

examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

GPCRX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors,

20   containing a nucleic acid encoding an GPCRX protein, or derivatives, fragments, analogs or

homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable

of transporting another nucleic acid to which it has been linked. One type of vector is a

"plasmid", which refers to a circular double stranded DNA loop into which additional DNA

segments can be ligated. Another type of vector is a viral vector, wherein additional DNA

25   segments can be ligated into the viral genome. Certain vectors are capable of autonomous

replication in a host cell into which they are introduced (e.g., bacterial vectors having a

bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g.,

non-episomal mammalian vectors) are integrated into the genome of a host cell upon

introduction into the host cell, and thereby are replicated along with the host genome.

30   Moreover, certain vectors are capable of directing the expression of genes to which they are

operatively-linked. Such vectors are referred to herein as "expression vectors". In general,

expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

In the present specification, "plasmid" and "vector" can be used interchangeably as the

plasmid is the most commonly used form of vector. However, the invention is intended to

include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that

5   the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in*

10   *vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN

15   ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be

20   transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, GPCRX proteins, mutant forms of GPCRX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of

25   GPCRX proteins in prokaryotic or eukaryotic cells. For example, GPCRX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and

30   translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors

101

typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the

5   recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse

10   glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)

15   60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.*, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the

20   nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the GPCRX expression vector is a yeast expression vector.

25   Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, GPCRX can be expressed in insect cells using baculovirus expression

30   vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.,* SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors

include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable

5   expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*,

10   tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.*

15   8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European

20   Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That

25   is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to GPCRX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory

30   sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene

expression using antisense genes *see, e.g.,* Weintraub, *et al.,* "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics,* Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, GPCRX protein can be expressed in bacterial cells such as *E. coli,* insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.,* DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.,* resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding GPCRX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.,* cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.,* express) GPCRX protein. Accordingly, the invention further provides

methods for producing GPCRX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding GPCRX protein has been introduced) in a suitable medium such that GPCRX protein is produced. In another embodiment, the method further

5 comprises isolating GPCRX protein from the medium or the host cell.

Transgenic GPCRX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which GPCRX protein-coding sequences have been introduced.

10 Such host cells can then be used to create non-human transgenic animals in which exogenous GPCRX sequences have been introduced into their genome or homologous recombinant animals in which endogenous GPCRX sequences have been altered. Such animals are useful for studying the function and/or activity of GPCRX protein and for identifying and/or evaluating modulators of GPCRX protein activity. As used herein, a "transgenic animal" is a

15 non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature

20 animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous GPCRX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell

25 of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing GPCRX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human GPCRX cDNA sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24,

30 26 and 27 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human GPCRX gene, such as a mouse GPCRX gene, can be isolated based on hybridization to the human GPCRX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be

included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the GPCRX transgene to direct expression of GPCRX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have 5   become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the GPCRX transgene in its genome and/or expression of GPCRX 10   mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding GPCRX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at 15   least a portion of an GPCRX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the GPCRX gene. The GPCRX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27), but more preferably, is a non-human homologue of a human GPCRX gene. For example, a mouse homologue of human GPCRX gene of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 20   17, 19, 20, 22, 24, 26 and 27 can be used to construct a homologous recombination vector suitable for altering an endogenous GPCRX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous GPCRX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

25   Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous GPCRX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous GPCRX protein). In the homologous recombination vector, the altered portion of the GPCRX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the 30   GPCRX gene to allow for homologous recombination to occur between the exogenous GPCRX gene carried by the vector and an endogenous GPCRX gene in an embryonic stem cell. The additional flanking GPCRX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et*

al., 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced GPCRX gene has homologously-recombined with the endogenous GPCRX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

5   The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in

10   their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

15   In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of

20   *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the

25   other containing a transgene encoding a recombinase.

          Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of

30   electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The GPCRX nucleic acid molecules, GPCRX proteins, and anti-GPCRX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (i.e., topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration,

suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating

5   action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of

10 surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by

15 including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., an GPCRX protein or anti-GPCRX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered

20 sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered

25 solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier

30 for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient

109

such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a

lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a

sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint,

methyl salicylate, or orange flavoring.

5   For administration by inhalation, the compounds are delivered in the form of an

aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g.,

a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For

transmucosal or transdermal administration, penetrants appropriate to the barrier to be

10   permeated are used in the formulation. Such penetrants are generally known in the art, and

include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid

derivatives. Transmucosal administration can be accomplished through the use of nasal sprays

or suppositories. For transdermal administration, the active compounds are formulated into

ointments, salves, gels, or creams as generally known in the art.

15   The compounds can also be prepared in the form of suppositories (e.g., with

conventional suppository bases such as cocoa butter and other glycerides) or retention enemas

for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect

the compound against rapid elimination from the body, such as a controlled release

20   formulation, including implants and microencapsulated delivery systems. Biodegradable,

biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides,

polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of

such formulations will be apparent to those skilled in the art. The materials can also be

obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal

25   suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral

antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared

according to methods known to those skilled in the art, for example, as described in U.S.

Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage

30   unit form for ease of administration and uniformity of dosage. Dosage unit form as used

herein refers to physically discrete units suited as unitary dosages for the subject to be treated;

each unit containing a predetermined quantity of active compound calculated to produce the

desired therapeutic effect in association with the required pharmaceutical carrier. The

specification for the dosage unit forms of the invention are dictated by and directly dependent

on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

5   The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.,* Chen, *et al.,* 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery

10   vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.,* retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

15   **Screening and Detection Methods**

The isolated nucleic acid molecules of the invention can be used to express GPCRX protein (*e.g.,* via a recombinant expression vector in a host cell in gene therapy applications), to detect GPCRX mRNA (*e.g.,* in a biological sample) or a genetic lesion in an GPCRX gene, and to modulate GPCRX activity, as described further, below. In addition, the GPCRX

20   proteins can be used to screen drugs or compounds that modulate the GPCRX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of GPCRX protein or production of GPCRX protein forms that have decreased or aberrant activity compared to GPCRX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the

25   metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-GPCRX antibodies of the invention can be used to detect and isolate GPCRX proteins and modulate GPCRX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the

30   disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra.*

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to GPCRX proteins or have a

5 stimulatory or inhibitory effect on, *e.g.*, GPCRX protein expression or GPCRX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an

10 GPCRX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity

15 chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small

20 molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art,

25 for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

30 Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science*

249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991.
J. Mol. Biol. 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a
membrane-bound form of GPCRX protein, or a biologically-active portion thereof, on the cell
5   surface is contacted with a test compound and the ability of the test compound to bind to an
GPCRX protein determined. The cell, for example, can of mammalian origin or a yeast cell.
Determining the ability of the test compound to bind to the GPCRX protein can be
accomplished, for example, by coupling the test compound with a radioisotope or enzymatic
label such that binding of the test compound to the GPCRX protein or biologically-active
10   portion thereof can be determined by detecting the labeled compound in a complex. For
example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly,
and the radioisotope detected by direct counting of radioemission or by scintillation counting.
Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish
peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by
15   determination of conversion of an appropriate substrate to product. In one embodiment, the
assay comprises contacting a cell which expresses a membrane-bound form of GPCRX
protein, or a biologically-active portion thereof, on the cell surface with a known compound
which binds GPCRX to form an assay mixture, contacting the assay mixture with a test
compound, and determining the ability of the test compound to interact with an GPCRX
20   protein, wherein determining the ability of the test compound to interact with an GPCRX
protein comprises determining the ability of the test compound to preferentially bind to
GPCRX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell
expressing a membrane-bound form of GPCRX protein, or a biologically-active portion
25   thereof, on the cell surface with a test compound and determining the ability of the test
compound to modulate (*e.g.*, stimulate or inhibit) the activity of the GPCRX protein or
biologically-active portion thereof. Determining the ability of the test compound to modulate
the activity of GPCRX or a biologically-active portion thereof can be accomplished, for
example, by determining the ability of the GPCRX protein to bind to or interact with an
30   GPCRX target molecule. As used herein, a "target molecule" is a molecule with which an
GPCRX protein binds or interacts in nature, for example, a molecule on the surface of a cell
which expresses an GPCRX interacting protein, a molecule on the surface of a second cell, a
molecule in the extracellular milieu, a molecule associated with the internal surface of a cell
membrane or a cytoplasmic molecule. An GPCRX target molecule can be a non-GPCRX

molecule or an GPCRX protein or polypeptide of the invention. In one embodiment, an

GPCRX target molecule is a component of a signal transduction pathway that facilitates

transduction of an extracellular signal (e.g. a signal generated by binding of a compound to a

membrane-bound GPCRX molecule) through the cell membrane and into the cell. The target,

5   for example, can be a second intercellular protein that has catalytic activity or a protein that

facilitates the association of downstream signaling molecules with GPCRX.

Determining the ability of the GPCRX protein to bind to or interact with an GPCRX

target molecule can be accomplished by one of the methods described above for determining

direct binding. In one embodiment, determining the ability of the GPCRX protein to bind to or

10   interact with an GPCRX target molecule can be accomplished by determining the activity of

the target molecule. For example, the activity of the target molecule can be determined by

detecting induction of a cellular second messenger of the target (i.e. intracellular Ca^{2+},

diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate

substrate, detecting the induction of a reporter gene (comprising an GPCRX-responsive

15   regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g.,

luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation,

or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising

contacting an GPCRX protein or biologically-active portion thereof with a test compound and

20   determining the ability of the test compound to bind to the GPCRX protein or biologically-

active portion thereof. Binding of the test compound to the GPCRX protein can be determined

either directly or indirectly as described above. In one such embodiment, the assay comprises

contacting the GPCRX protein or biologically-active portion thereof with a known compound

which binds GPCRX to form an assay mixture, contacting the assay mixture with a test

25   compound, and determining the ability of the test compound to interact with an GPCRX

protein, wherein determining the ability of the test compound to interact with an GPCRX

protein comprises determining the ability of the test compound to preferentially bind to

GPCRX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting

30   GPCRX protein or biologically-active portion thereof with a test compound and determining

the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the

GPCRX protein or biologically-active portion thereof. Determining the ability of the test

compound to modulate the activity of GPCRX can be accomplished, for example, by

determining the ability of the GPCRX protein to bind to an GPCRX target molecule by one of

the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of GPCRX protein can be accomplished by determining the ability of the GPCRX protein further modulate an GPCRX target molecule. For example, the catalytic/enzymatic activity of the target molecule

5   on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the GPCRX protein or biologically-active portion thereof with a known compound which binds GPCRX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an GPCRX protein, wherein

10   determining the ability of the test compound to interact with an GPCRX protein comprises determining the ability of the GPCRX protein to preferentially bind to or modulate the activity of an GPCRX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of GPCRX protein. In the case of cell-free assays comprising the

15   membrane-bound form of GPCRX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of GPCRX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®,

20   Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either GPCRX protein or its target molecule to facilitate separation of

25   complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to GPCRX protein, or interaction of GPCRX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a

30   fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-GPCRX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or GPCRX protein, and the mixture

is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*.

5   Alternatively, the complexes can be dissociated from the matrix, and the level of GPCRX protein binding or activity determined using standard techniques.

        Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the GPCRX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated

10   GPCRX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with GPCRX protein or target molecules, but which do not interfere with binding of the GPCRX protein to its target

15   molecule, can be derivatized to the wells of the plate, and unbound target or GPCRX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the GPCRX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity

20   associated with the GPCRX protein or target molecule.

        In another embodiment, modulators of GPCRX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of GPCRX mRNA or protein in the cell is determined. The level of expression of GPCRX mRNA or protein in the presence of the candidate compound is compared to the level of expression of

25   GPCRX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of GPCRX mRNA or protein expression based upon this comparison. For example, when expression of GPCRX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of GPCRX mRNA or

30   protein expression. Alternatively, when expression of GPCRX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of GPCRX mRNA or protein expression. The level of GPCRX mRNA or protein expression in the cells can be determined by methods described herein for detecting GPCRX mRNA or protein.

In yet another aspect of the invention, the GPCRX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8:

5   1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with GPCRX ("GPCRX-binding proteins" or "GPCRX-bp") and modulate GPCRX activity. Such GPCRX-binding proteins are also likely to be involved in the propagation of signals by the GPCRX proteins as, for example, upstream or downstream elements of the GPCRX pathway.

The two-hybrid system is based on the modular nature of most transcription factors,

10   which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for GPCRX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation

15   domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an GPCRX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be

20   detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with GPCRX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

25   Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic

30   identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the GPCRX sequences, SEQ ID

5   NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or fragments or derivatives thereof, can be used to map the location of the GPCRX genes, respectively, on a chromosome. The mapping of the GPCRX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, GPCRX genes can be mapped to chromosomes by preparing PCR primers

10   (preferably 15-25 bp in length) from the GPCRX sequences. Computer analysis of the GPCRX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the GPCRX sequences will

15   yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in

20   which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, et al.,

25   1983. Science 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using

30   a single thermal cycler. Using the GPCRX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in

metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases.

5   However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

10   Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations

15   during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes

20   and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the GPCRX gene, can be determined. If a mutation

25   is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete

30   sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The GPCRX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with

119

one or more restriction enzymes, and probed on a Southern blot to yield unique bands for

identification. The sequences of the invention are useful as additional DNA markers for RFLP

("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative

5   technique that determines the actual base-by-base DNA sequence of selected portions of an

individual's genome. Thus, the GPCRX sequences described herein can be used to prepare

two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used

to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner,

10   can provide unique individual identifications, as each individual will have a unique set of such

DNA sequences due to allelic differences. The sequences of the invention can be used to

obtain such identification sequences from individuals and from tissue. The GPCRX sequences

of the invention uniquely represent portions of the human genome. Allelic variation occurs to

some degree in the coding regions of these sequences, and to a greater degree in the noncoding

15   regions. It is estimated that allelic variation between individual humans occurs with a

frequency of about once per each 500 bases. Much of the allelic variation is due to single

nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms

(RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard

20   against which DNA from an individual can be compared for identification purposes. Because

greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are

necessary to differentiate individuals. The noncoding sequences can comfortably provide

positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a

noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in

25   SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27 are used, a more

appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic

assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for

30   prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly,

one aspect of the invention relates to diagnostic assays for determining GPCRX protein and/or

nucleic acid expression as well as GPCRX activity, in the context of a biological sample (e.g.,

blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a

disease or disorder, or is at risk of developing a disorder, associated with aberrant GPCRX

expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic

5   syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with GPCRX protein, nucleic acid expression or activity. For example, mutations in an GPCRX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby

10   prophylactically treat an individual prior to the onset of a disorder characterized by or associated with GPCRX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining GPCRX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics").

15   Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*,

20   drugs, compounds) on the expression or activity of GPCRX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of GPCRX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological

25   sample with a compound or an agent capable of detecting GPCRX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes GPCRX protein such that the presence of GPCRX is detected in the biological sample. An agent for detecting GPCRX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to GPCRX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length GPCRX nucleic acid, such as the

30   nucleic acid of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to GPCRX mRNA or

genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting GPCRX protein is an antibody capable of binding to GPCRX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect GPCRX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of GPCRX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of GPCRX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of GPCRX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of GPCRX protein include introducing into a subject a labeled anti-GPCRX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting GPCRX protein, mRNA, or genomic DNA, such that the presence of GPCRX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of GPCRX protein, mRNA or genomic DNA in the control sample with the presence of GPCRX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of GPCRX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting GPCRX protein or mRNA in a biological sample; means for determining the amount of GPCRX in the sample; and means for comparing the amount of GPCRX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCRX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant GPCRX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with GPCRX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant GPCRX expression or activity in which a test sample is obtained from a subject and GPCRX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of GPCRX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant GPCRX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant GPCRX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant GPCRX expression or activity in which a test sample is obtained and GPCRX protein or nucleic acid is detected (*e.g.*, wherein the presence of GPCRX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant GPCRX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an GPCRX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a

5   genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an GPCRX-protein, or the misexpression of the GPCRX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from an GPCRX gene; (*ii*) an addition of one or more nucleotides to an GPCRX gene; (*iii*) a substitution of one or more nucleotides of an GPCRX gene, (*iv*) a

10   chromosomal rearrangement of an GPCRX gene; (*v*) an alteration in the level of a messenger RNA transcript of an GPCRX gene, (*vi*) aberrant modification of an GPCRX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an GPCRX gene, (*viii*) a non-wild-type level of an GPCRX protein, (*ix*) allelic loss of an GPCRX gene, and (*x*) inappropriate post-translational

15   modification of an GPCRX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an GPCRX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

20   In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point

25   mutations in the GPCRX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an GPCRX gene under conditions such that hybridization and amplification of the GPCRX gene (if present)

30   occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.,* 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al,* 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification

5   method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an GPCRX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and

10   control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or

15   loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in GPCRX can be identified by hybridizing a sample and control nucleic acids, *e.g.,* DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.,* 1996. *Human Mutation* 7: 244-255; Kozal, *et al.,* 1996. *Nat. Med.* 2: 753-759. For example, genetic

20   mutations in GPCRX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al., supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second

25   hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art

30   can be used to directly sequence the GPCRX gene and detect mutations by comparing the sequence of the sample GPCRX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures

can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.,* 1995.

Biotechniques 19: 448), including sequencing by mass spectrometry (see, *e.g.,* PCT

International Publication No. WO 94/16101; Cohen, *et al.,* 1996. *Adv. Chromatography* 36:

127-162; and Griffin, *et al.,* 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

5   Other methods for detecting mutations in the GPCRX gene include methods in which

protection from cleavage agents is used to detect mismatched bases in RNA/RNA or

RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.,* 1985. *Science* 230: 1242. In general, the

art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by

hybridizing (labeled) RNA or DNA containing the wild-type GPCRX sequence with

10   potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded

duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as

which will exist due to basepair mismatches between the control and sample strands. For

instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with

S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either

15   DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide

and with piperidine in order to digest mismatched regions. After digestion of the mismatched

regions, the resulting material is then separated by size on denaturing polyacrylamide gels to

determine the site of mutation. *See, e.g.,* Cotton, *et al.,* 1988. *Proc. Natl. Acad. Sci. USA* 85:

4397; Saleeba, *et al.,* 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control

20   DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more

proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA

mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in

GPCRX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli*

25   cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T

at G/T mismatches. *See, e.g.,* Hsu, *et al.,* 1994. *Carcinogenesis* 15: 1657-1662. According to

an exemplary embodiment, a probe based on an GPCRX sequence, *e.g.,* a wild-type GPCRX

sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is

treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be

30   detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify

mutations in GPCRX genes. For example, single strand conformation polymorphism (SSCP)

may be used to detect differences in electrophoretic mobility between mutant and wild type

nucleic acids. *See, e.g.,* Orita, *et al.,* 1989. *Proc. Natl. Acad. Sci. USA:* 86: 2766; Cotton,

1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79.

Single-stranded DNA fragments of sample and control GPCRX nucleic acids will be

denatured and allowed to renature. The secondary structure of single-stranded nucleic acids

varies according to sequence, the resulting alteration in electrophoretic mobility enables the

5   detection of even a single base change. The DNA fragments may be labeled or detected with

labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than

DNA), in which the secondary structure is more sensitive to a change in sequence. In one

embodiment, the subject method utilizes heteroduplex analysis to separate double stranded

heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et*

10   *al.*, 1991. *Trends Genet.* 7: 5.

        In yet another embodiment, the movement of mutant or wild-type fragments in

polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient

gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is

used as the method of analysis, DNA will be modified to insure that it does not completely

15   denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich

DNA by PCR. In a further embodiment, a temperature gradient is used in place of a

denaturing gradient to identify differences in the mobility of control and sample DNA. *See,*

e.g., Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

        Examples of other techniques for detecting point mutations include, but are not limited

20   to, selective oligonucleotide hybridization, selective amplification, or selective primer

extension. For example, oligonucleotide primers may be prepared in which the known

mutation is placed centrally and then hybridized to target DNA under conditions that permit

hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163;

Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides

25   are hybridized to PCR amplified target DNA or a number of different mutations when the

oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target

DNA.

        Alternatively, allele specific amplification technology that depends on selective PCR

amplification may be used in conjunction with the instant invention. Oligonucleotides used as

30   primers for specific amplification may carry the mutation of interest in the center of the

molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*,

1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where,

under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,*

Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel

restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a

5   perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.,* in clinical settings to diagnose

10   patients exhibiting symptoms or family history of a disease or illness involving an GPCRX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which GPCRX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal

15   mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on GPCRX activity (*e.g.,* GPCRX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The

20   disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such

25   treatment, the pharmacogenomics (*i.e.,* the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the

30   selection of effective agents (*e.g.,* drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of GPCRX protein, expression of GPCRX nucleic acid, or mutation content of GPCRX genes in an

individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of GPCRX protein, expression of GPCRX nucleic acid, or mutation content of GPCRX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness

phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an GPCRX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

5   **Monitoring of Effects During Clinical Trials**

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of GPCRX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to

10   increase GPCRX gene expression, protein levels, or upregulate GPCRX activity, can be monitored in clinical trails of subjects exhibiting decreased GPCRX gene expression, protein levels, or downregulated GPCRX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease GPCRX gene expression, protein levels, or downregulate GPCRX activity, can be monitored in clinical trails of subjects exhibiting

15   increased GPCRX gene expression, protein levels, or upregulated GPCRX activity. In such clinical trials, the expression or activity of GPCRX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including GPCRX, that are

20   modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates GPCRX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of GPCRX and other genes implicated in the disorder. The levels of gene

25   expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of GPCRX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be

30   determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the

130

screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration

sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression

of an GPCRX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*)

obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of

5   expression or activity of the GPCRX protein, mRNA, or genomic DNA in the

post-administration samples; (*v*) comparing the level of expression or activity of the GPCRX

protein, mRNA, or genomic DNA in the pre-administration sample with the GPCRX protein,

mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the

administration of the agent to the subject accordingly. For example, increased administration

10   of the agent may be desirable to increase the expression or activity of GPCRX to higher levels

than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased

administration of the agent may be desirable to decrease expression or activity of GPCRX to

lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

15   The invention provides for both prophylactic and therapeutic methods of treating a

subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant

GPCRX expression or activity. The disorders include cardiomyopathy, atherosclerosis,

hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD),

atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis,

20   ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity,

transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer,

neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation,

idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS,

bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary

25   Ostoeodystrophy, and other diseases, disorders and conditions of the like.

            These methods of treatment will be discussed more fully, below.

Disease and Disorders

            Diseases and disorders that are characterized by increased (relative to a subject not

suffering from the disease or disorder) levels or biological activity may be treated with

30   Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize

activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may

be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs,

derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*)

131

nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (i.e., due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous

5   recombination (see, e.g., Capecchi, 1989. Science 244: 1288-1292); or (v) modulators (i.e., inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

          Diseases and disorders that are characterized by decreased (relative to a subject not

10   suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (i.e., are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

15   Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (e.g., from biopsy tissue) and assaying it in vitro for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (e.g., by Western blot analysis, immunoprecipitation followed by

20   sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (e.g., Northern assays, dot blots, in situ hybridization, and the like).

Prophylactic Methods

         In one aspect, the invention provides a method for preventing, in a subject, a disease or

25   condition associated with an aberrant GPCRX expression or activity, by administering to the subject an agent that modulates GPCRX expression or at least one GPCRX activity. Subjects at risk for a disease that is caused or contributed to by aberrant GPCRX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation

30   of symptoms characteristic of the GPCRX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of GPCRX aberrancy, for example, an GPCRX agonist or GPCRX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays

described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating GPCRX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of GPCRX protein activity associated with the cell. An agent that modulates GPCRX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an GPCRX protein, a peptide, an GPCRX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more GPCRX protein activity. Examples of such stimulatory agents include active GPCRX protein and a nucleic acid molecule encoding GPCRX that has been introduced into the cell. In another embodiment, the agent inhibits one or more GPCRX protein activity. Examples of such inhibitory agents include antisense GPCRX nucleic acid molecules and anti-GPCRX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an GPCRX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) GPCRX expression or activity. In another embodiment, the method involves administering an GPCRX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant GPCRX expression or activity.

Stimulation of GPCRX activity is desirable in situations in which GPCRX is abnormally downregulated and/or in which increased GPCRX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

133

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows,

5 monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The GPCRX nucleic acids and proteins of the invention are useful in potential

10 prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders

15 associated with chronic diseases and various cancers.

As an example, a cDNA encoding the GPCRX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease,

20 anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the GPCRX protein, and the GPCRX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the

25 presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.,* some peptides have been found to possess anti-bacterial properties). These materials are further useful in the generation of antibodies which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

30 **Examples**

134

Example 1. Identification of GPCRX clones

The novel nucleic acid sequences of GPCR1 through GPCR5 were identified on chromosome 11 by TblastN using CuraGen Corporation's sequence files for Olfactory Receptor homolog, run against the Genomic Daily Files made available by GenBank. The 165 kbp human genomic clone from CuraGen acc:AP0010804HTG derived from Homo sapiens chromosome 11, clone RP11-164A10 map 11q, was analyzed by GenScan and Grail software to identify exons and putative coding sequences. These clones were also analyzed by TblastN, BlastX and other programs to identify genomic regions translating to proteins with similarity to the original protein or protein family of interest.

All novel GPCRX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

135

Example 2. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed

5   on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells

10   ` and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to constitutively expressed genes such as β-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog

15   No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48^0C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to

20   the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50^0C; 10 min. at 95^0C; 15 sec. at 95^0C/1 min. at 60^0C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta

25   CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their β-actin /GAPDH average CT values.

30   Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software

package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum
5   primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations
10   were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two
.   probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems
15   7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/μl RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

20   In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,
* = established from metastasis,
met = metastasis,
s cell var= small cell variant,
25   non-s = non-sm =non-small,
squam = squamous,
pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
30   neuro = neuroblastoma.

Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close
35   cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins"

obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be

indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

5   Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was

10   obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

          Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells,

15   microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF

20   alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

          Mononuclear cells were prepared from blood of employees at CuraGen Corporation,

25   using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 μg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at

30   5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μg/ml. Samples

were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction)·
samples were obtained by taking blood from two donors, isolating the mononuclear cells using
Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately
2×10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1

5   mM sodium pyruvate (Gibco), mercaptoethanol (5.5×10^{-5} M) (Gibco), and 10 mM Hepes
(Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7
days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS
selection columns and a Vario Magnet according to the manufacturer's instructions.

10   Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum
(FCS) (Hyclone, Logan, UT), 100 μM non essential amino acids (Gibco), 1 mM sodium
pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml
GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes
for 5-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM

15   sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco) and
10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and
dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100
ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody
(Pharmingen) at 10 μg/ml for 6 and 12-14 hours.

20   CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from
mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns
and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4
lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19
cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then

25   CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells
being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes
were placed in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM
sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco)
and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated

30   overnight with 0.5 μg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in
PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare
chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days
on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in

DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second

5   activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile

10 dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resuspended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 μg/ml or anti-CD40 (Pharmingen) at approximately 10 μg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at

15 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μg/ml anti-CD28 (Pharmingen) and 2 μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5 -10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM

20 non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 μg/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 μg/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7

25 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μg/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes

30 were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and

third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5×10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 $\times 10^5$ cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 μl of RNAse-free water and 35 μl buffer (Promega) 5 μl DTT, 7 μl RNAsin and 8 μl DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C.

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor.

All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's

5   disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by

10   neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

15   RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe

20   and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

25   Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

30   **A. GPCR1 (also known as AP001804_A or CG54326-01)**

Expression of gene AP001804_A was assessed using the primer-probe sets Ag1634 and Ag2357 (identical sequences), described in Table 12. Results of the RTQ-PCR runs are shown in Tables 13 and 14.

143

Table 12. Probe Name Ag1634/Ag2357

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TGAACTTTGTTCCAGAGGAGAA-3'	59	22	248	80
Probe	TET-5'-TCTCCTTTCTGGAATGCATTACTCAA-3'-TAMRA	64.3	26	275	81
Reverse	5'-GGTAGCCTTCTGCAATTACAAA-3'	58.5	22	319	82

Table 13. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5589 _ag1634_b2	Tissue Name	Relative Expression(%) 1.3dx4tm5589 _ag1634_b2
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	4.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	6.5
Brain (thalamus)	6.1	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	4.7	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	9.3
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	4.9
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.0	Mammary gland	0.0
CNS ca. (glio) U251	0.0	Breast ca.* (pl. effusion) MCF-7	28.7
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	37.1
Heart	0.0	Breast ca. BT-549	0.0

Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	3.7	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	4.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	**100.0**	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	5.2
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	4.7
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	4.5	Prostate ca.* (bone met)PC-3	4.2
Colon ca. HCT-116	0.0	Testis	9.3
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	23.7
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table 14. Panel 4D

Tissue Name	Relative Expression(%) 4dx4tm5519t_ ag1634_a2	Tissue Name	Relative Expression(%) 4dx4tm5519t_ ag1634_a2
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	93781_HUVEC (Endothelial)_IL-11	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	93583_Lung Microvascular Endothelial Cells_none	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml)	0.0

93565_primary Th1_resting dy 4-6 in IL-2	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0
93354_CD4_none	0.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0
93103_LAK cells_resting	0.0
93788_LAK cells_IL-2	0.0
93787_LAK cells_IL-2+IL-12	0.0
93789_LAK cells_IL-2+IFN gamma	0.0
93790_LAK cells_IL-2+ IL-18	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0
93578_NK Cells IL-2_resting	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0
93249_Ramos (B cell)_none	0.0

and IL1b (1 ng/ml)	
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0
93347_Small Airway Epithelium_none	0.0
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
92668_Coronery Artery SMC_resting	0.0
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93107_astrocytes_resting	0.0
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
92666_KU-812 (Basophil)_resting	0.2
92667_KU-812 (Basophil)_PMA/ionoycin	0.4
93579_CCD1106 (Keratinocytes)_none	0.0
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0
93791_Liver Cirrhosis	5.3
93792_Lupus Kidney	0.0
93577_NCI-H292	0.0
93358_NCI-H292_IL-4	0.0
93360_NCI-H292_IL-9	0.0
93359_NCI-H292_IL-13	0.0
93357_NCI-H292_IFN gamma	0.0
93777_HPAEC_-	0.0
93778_HPAEC_IL-1 beta/TNA alpha	0.5
93254_Normal Human Lung Fibroblast_none	0.2
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93257_Normal Human Lung Fibroblast_IL-4	0.4
93256_Normal Human Lung	0.7

		Fibroblast_IL-9	
93250_Ramos (B cell)_ionomycin	0.0	93255_Normal Human Lung Fibroblast_IL-13	0.5
93349_B lymphocytes_PWM	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0
93350_B lymphoytes_CD40L and IL-4	0.0	93106_Dermal Fibroblasts CCD1070_resting	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAion omycin	0.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0
93356_Dendritic Cells_none	0.2	93772_dermal fibroblast_IFN gamma	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	93771_dermal fibroblast_IL-4	0.0
93775_Dendritic Cells_anti-CD40	4.2	93259_IBD Colitis 1**	100.0
93774_Monocytes_resting	0.0	93260_IBD Colitis 2	0.8
93776_Monocytes_LPS 50 ng/ml	0.0	93261_IBD Crohns	0.0
93581_Macrophages_resting	2.7	735010_Colon_normal	7.3
93582_Macrophages_LPS 100 ng/ml	0.0	735019_Lung_none	0.4
93098_HUVEC (Endothelial)_none	0.0	64028-1_Thymus_none	0.0
93099_HUVEC (Endothelial)_starved	0.0	64030-1_Kidney_none	0.0

Panel 1.3D Summary: Ag1634 Expression of GPCR1 gene AP001804_A is low/undetectable (CT values >35) in all cell lines and tissues except for spleen. Therefore, this gene may be used to distinguish spleen from other tissues. Ag2357 Expression was low/undetectable (CT values 40) in all tissues tested.

Panel 2D Summary: Ag2357 Expression was low/undetectable (CT values 40) in all tissues tested and thus the results not shown.

Panel 2.2 Summary: Ag1634 Expression of gene AP001804_A is low/undetectable (CT values >35) in all cell lines and tissues on this panel thus the results not shown.

Panel 4D Summary: Ag1634 Expression of the AP001804_A transcript is detected in colitis 1 and in dendritic cells treated with anti-CD40. The protein encoded for by this antigen may be important in the inflammatory process and particularly in the function of activated dendritic cells. Antagonistic antibodies or small molecule therapeutics that inhibit AP001804_A protein function may therefore reduce or inhibit inflammation in the bowel due to inflammatory bowl disease (IBD). Ag2357 Expression was low/undetectable (CT values 40) in all tissues tested and chemistry control did not work well (CT = 35).

147

B. GPCR2 (also known as AP001804_B or CG54335-01)

Expression of gene AP001804_B was assessed using the primer-probe sets Ag2355 and Ag1635 (identical sequences), described in Table 15. Results of the RTQ-PCR runs are shown in Tables 16, 17, 18, and 19.

5

Table 15. Probe Name Ag2355/Ag1635

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TCATACAAGTGCCATGATGAAA-3'	59	22	474	83
Probe	FAM-5'-TGTCCTTTTGCAAATCCCACATTATCA-3'-TAMRA	68	27	497	84
Reverse	5'-AGGGGAAGAACATCACAGAAGT-3'	59.1	22	530	85

Table 16. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5627f _ag2355_b1	Tissue Name	Relative Expression(%) 1.3dx4tm5627f _ag2355_b1
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
CNS ca. (glio/astro) U87-MG	8.1	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	7.1
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	7.7
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0

Tissue Name	Value	Tissue Name	Value
CNS ca. (glio) SNB-19	6.4	Mammary gland	0.0
CNS ca. (glio) U251	0.0	Breast ca.* (pl. effusion) MCF-7	100.0
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	57.3
Heart	0.0	Breast ca. BT-549	0.0
Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	6.1	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	19.2
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	4.7
Colon ca. HCT-116	0.0	Testis	53.1
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	10.9
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table 17. Panel 2D

Tissue Name	Relative Expression(%) 2dx4tm4923f_ag2355_a2	Tissue Name	Relative Expression(%) 2dx4tm4923f_ag2355_a2
Normal Colon GENPAK 061003	4.3	Kidney NAT Clontech 8120608	0.0
83219 CC Well to Mod Diff (ODO3866)	0.7	Kidney Cancer Clontech 8120613	0.0
83220 CC NAT (ODO3866)	3.7	Kidney NAT Clontech 8120614	0.4
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	Kidney Cancer Clontech 9010320	0.4
83222 CC NAT (ODO3868)	0.0	Kidney NAT Clontech 9010321	0.0
83235 CC Mod Diff (ODO3920)	0.0	Normal Uterus GENPAK 061018	1.0
83236 CC NAT (ODO3920)	0.3	Uterus Cancer GENPAK	0.0

83237 CC Gr.2 ascend colon (ODO3921)	0.3	064011 Normal Thyroid Clontech A+ 6570-1	0.0
83238 CC NAT (ODO3921)	1.4	Thyroid Cancer GENPAK 064010	0.0
83241 CC from Partial Hepatectomy (ODO4309)	0.0	Thyroid Cancer INVITROGEN A302152	0.0
83242 Liver NAT (ODO4309)	0.0	Thyroid NAT INVITROGEN A302153	0.0
87472 Colon mets to lung (OD04451-01)	0.0	Normal Breast GENPAK 061019	0.0
87473 Lung NAT (OD04451-02)	0.8	84877 Breast Cancer (OD04566)	0.0
Normal Prostate Clontech A+ 6546-1	0.8	85975 Breast Cancer (OD04590-01)	0.0
84140 Prostate Cancer (OD04410)	0.0	85976 Breast Cancer Mets (OD04590-03)	0.0
84141 Prostate NAT (OD04410)	0.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0
87073 Prostate Cancer (OD04720-01)	0.0	GENPAK Breast Cancer 064006	0.0
87074 Prostate NAT (OD04720-02)	0.5	Breast Cancer Res. Gen. 1024	0.0
Normal Lung GENPAK 061010	1.8	Breast Cancer Clontech 9100266	0.0
83239 Lung Met to Muscle (ODO4286)	0.8	Breast NAT Clontech 9100265	0.4
83240 Muscle NAT (ODO4286)	0.0	Breast Cancer INVITROGEN A209073	0.0
84136 Lung Malignant Cancer (OD03126)	0.0	Breast NAT INVITROGEN A2090734	0.4
84137 Lung NAT (OD03126)	0.0	Normal Liver GENPAK 061009	0.0
84871 Lung Cancer (OD04404)	0.0	Liver Cancer GENPAK 064003	5.0
84872 Lung NAT (OD04404)	0.0	Liver Cancer Research Genetics RNA 1025	0.7
84875 Lung Cancer (OD04565)	0.0	Liver Cancer Research Genetics RNA 1026	0.0
84876 Lung NAT (OD04565)	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.3
85950 Lung Cancer (OD04237-01)	0.3	Paired Liver Tissue Research Genetics RNA 6004-N	0.7
85970 Lung NAT (OD04237-02)	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0
83256 Liver NAT (ODO4310)	0.0	Normal Bladder GENPAK 061001	1.8
84139 Melanoma Mets to Lung (OD04321)	0.0	Bladder Cancer Research Genetics RNA 1023	0.0
84138 Lung NAT (OD04321)	0.0	Bladder Cancer INVITROGEN	18.1

		A302173	
Normal Kidney GENPAK 061008	0.9	87071 Bladder Cancer (OD04718-01)	0.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	2.1	87072 Bladder Normal Adjacent (OD04718-03)	0.0
83787 Kidney NAT (OD04338)	0.0	Normal Ovary Res. Gen.	0.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.8	Ovarian Cancer GENPAK 064008	0.0
83789 Kidney NAT (OD04339)	0.8	87492 Ovary Cancer (OD04768-07)	100.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	87493 Ovary NAT (OD04768-08)	0.0
83791 Kidney NAT (OD04340)	0.0	Normal Stomach GENPAK 061017	0.2
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer Clontech 9060358	0.0
83793 Kidney NAT (OD04348)	0.0	NAT Stomach Clontech 9060359	0.0
87474 Kidney Cancer (OD04622-01)	0.0	Gastric Cancer Clontech 9060395	0.0
87475 Kidney NAT (OD04622-03)	0.0	NAT Stomach Clontech 9060394	0.0
85973 Kidney Cancer (OD04450-01)	0.0	Gastric Cancer Clontech 9060397	0.0
85974 Kidney NAT (OD04450-03)	0.0	NAT Stomach Clontech 9060396	0.0
Kidney Cancer Clontech 8120607	0.0	Gastric Cancer GENPAK 064005	0.7

Table 18. Panel 3D

Tissue Name	Relative Expression(%) 3dx4tm5123f_ ag2355_a2	Tissue Name	Relative Expression(%) 3dx4tm5123f_ ag2355_a2
94905_Daoy_Medulloblastoma/ Cerebellum_sscDNA	0.0	94954_Ca Ski_Cervical epidermoid carcinoma (metastasis)_sscDNA	0.0
94906_TE671_Medulloblastom /Cerebellum_sscDNA	13.4	94955_ES-2_Ovarian clear cell carcinoma_sscDNA	0.0
94907_D283 Med_Medulloblastoma/Cerebell um_sscDNA	0.0	94957_Ramos/6h stim_"; Stimulated with PMA/ionomycin 6h_sscDNA	0.0
94908_PFSK-1_Primitive Neuroectodermal/Cerebellum_s scDNA	0.0	94958_Ramos/14h stim_"; Stimulated with PMA/ionomycin 14h_sscDNA	0.0
94909_XF-498_CNS_sscDNA	0.0	94962_MEG-01_Chronic myelogenous leukemia (megokaryoblast)_sscDNA	0.0
94910_SNB-78_CNS/glioma_sscDNA	0.0	94963_Raji_Burkitt's lymphoma_sscDNA	0.0

151

94911_SF-268_CNS/glioblastoma_sscDNA	0.0	94964_Daudi_Burkitt's lymphoma_sscDNA	0.0
94912_T98G_Glioblastoma_sscDNA	0.0	94965_U266_B-cell plasmacytoma/myeloma_sscDNA	72.8
96776_SK-N-SH_Neuroblastoma (metastasis)_sscDNA	0.0	94968_CA46_Burkitt's lymphoma_sscDNA	0.0
94913_SF-295_CNS/glioblastoma_sscDNA	0.0	94970_RL_non-Hodgkin's B-cell lymphoma_sscDNA	0.0
94914_Cerebellum_sscDNA	0.0	94972_JM1_pre-B-cell lymphoma/leukemia_sscDNA	0.0
96777_Cerebellum_sscDNA	0.0	94973_Jurkat_T cell leukemia_sscDNA	0.0
94916_NCI-H292_Mucoepidermoid lung carcinoma_sscDNA	0.0	94974_TF-1_Erythroleukemia_sscDNA	0.0
94917_DMS-114_Small cell lung cancer_sscDNA	57.8	94975_HUT 78_T-cell lymphoma_sscDNA	3.6
94918_DMS-79_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94977_U937_Histiocytic lymphoma_sscDNA	0.0
94919_NCI-H146_Small cell lung cancer/neuroendocrine_sscDNA	4.9	94980_KU-812_Myelogenous leukemia_sscDNA	0.0
94920_NCI-H526_Small cell lung cancer/neuroendocrine_sscDNA	4.1	94981_769-P_Clear cell renal carcinoma_sscDNA	0.0
94921_NCI-N417_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94983_Caki-2_Clear cell renal carcinoma_sscDNA	0.0
94923_NCI-H82_Small cell lung cancer/neuroendocrine_sscDNA	0.0	94984_SW 839_Clear cell renal carcinoma_sscDNA	0.0
94924_NCI-H157_Squamous cell lung cancer (metastasis)_sscDNA	28.9	94986_G401_Wilms' tumor_sscDNA	0.0
94925_NCI-H1155_Large cell lung cancer/neuroendocrine_sscDNA	0.0	94987_Hs766T_Pancreatic carcinoma (LN metastasis)_sscDNA	0.0
94926_NCI-H1299_Large cell lung cancer/neuroendocrine_sscDNA	100.0	94988_CAPAN-1_Pancreatic adenocarcinoma (liver metastasis)_sscDNA	0.0
94927_NCI-H727_Lung carcinoid_sscDNA	0.0	94989_SU86.86_Pancreatic carcinoma (liver metastasis)_sscDNA	0.0
94928_NCI-UMC-11_Lung carcinoid_sscDNA	47.1	94990_BxPC-3_Pancreatic adenocarcinoma_sscDNA	0.0
94929_LX-1_Small cell lung cancer_sscDNA	0.0	94991_HPAC_Pancreatic adenocarcinoma_sscDNA	0.0
94930_Colo-205_Colon cancer_sscDNA	0.0	94992_MIA PaCa-2_Pancreatic carcinoma_sscDNA	0.0

152

Tissue Name	Rel. Expr.	Tissue Name	Rel. Expr.
94931_KM12_Colon cancer_sscDNA	0.0	94993_CFPAC-1_Pancreatic ductal adenocarcinoma_sscDNA	0.0
94932_KM20L2_Colon cancer_sscDNA	0.0	94994_PANC-1_Pancreatic epithelioid ductal carcinoma_sscDNA	0.0
94933_NCI-H716_Colon cancer_sscDNA	0.0	94996_T24_Bladder carcinma (transitional cell)_sscDNA	0.0
94935_SW-48_Colon adenocarcinoma_sscDNA	0.0	94997_5637_Bladder carcinoma_sscDNA	0.0
94936_SW1116_Colon adenocarcinoma_sscDNA	0.0	94998_HT-1197_Bladder carcinoma_sscDNA	0.0
94937_LS 174T_Colon adenocarcinoma_sscDNA	0.0	94999_UM-UC-3_Bladder carcinma (transitional cell)_sscDNA	0.0
94938_SW-948_Colon adenocarcinoma_sscDNA	0.0	95000_A204_Rhabdomyosarcoma_sscDNA	0.0
94939_SW-480_Colon adenocarcinoma_sscDNA	0.0	95001_HT-1080_Fibrosarcoma_sscDNA	0.0
94940_NCI-SNU-5_Gastric carcinoma_sscDNA	0.0	95002_MG-63_Osteosarcoma (bone)_sscDNA	3.5
94941_KATO III_Gastric carcinoma_sscDNA	0.0	95003_SK-LMS-1_Leiomyosarcoma (vulva)_sscDNA	0.0
94943_NCI-SNU-16_Gastric carcinoma_sscDNA	0.0	95004_SJRH30_Rhabdomyosarcoma (met to bone marrow)_sscDNA	0.0
94944_NCI-SNU-1_Gastric carcinoma_sscDNA	0.0	95005_A431_Epidermoid carcinoma_sscDNA	0.0
94946_RF-1_Gastric adenocarcinoma_sscDNA	0.0	95007_WM266-4_Melanoma_sscDNA	5.6
94947_RF-48_Gastric adenocarcinoma_sscDNA	0.0	95010_DU 145_Prostate carcinoma (brain metastasis)_sscDNA	0.0
96778_MKN-45_Gastric carcinoma_sscDNA	0.0	95012_MDA-MB-468_Breast adenocarcinoma_sscDNA	0.0
94949_NCI-N87_Gastric carcinoma_sscDNA	6.9	95013_SCC-4_Squamous cell carcinoma of tongue_sscDNA	0.0
94951_OVCAR-5_Ovarian carcinoma_sscDNA	0.0	95014_SCC-9_Squamous cell carcinoma of tongue_sscDNA	0.0
94952_RL95-2_Uterine carcinoma_sscDNA	0.0	95015_SCC-15_Squamous cell carcinoma of tongue_sscDNA	0.0
94953_HelaS3_Cervical adenocarcinoma_sscDNA	2.7	95017_CAL 27_Squamous cell carcinoma of tongue_sscDNA	0.0

Table 19. Panel 4D

Tissue Name	Relative Expression(%) 4dx4tm4927f_ ag2355_a2	Tissue Name	Relative Expression(%) 4dx4tm4927f_ ag2355_a2

153

93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	93781_HUVEC (Endothelial)_IL-11	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	93583_Lung Microvascular Endothelial Cells_none	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	93347_Small Airway Epithelium_none	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	92668_Coronery Artery SMC_resting	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	14.7
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	93107_astrocytes_resting	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	92666_KU-812 (Basophil)_resting	0.0
93354_CD4_none	0.0	92667_KU-812 (Basophil)_PMA/ionoycin	22.3
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	93579_CCD1106 (Keratinocytes)_none	0.0
		93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0
93103_LAK cells_resting	0.0	93791_Liver Cirrhosis	71.5
93788_LAK cells_IL-2	0.0	93792_Lupus Kidney	0.0
93787_LAK cells_IL-2+IL-12	0.0	93577_NCI-H292	0.0
93789_LAK cells_IL-2+IFN	0.0		

154

gamma

93790_LAK cells_IL-2+ IL-18	0.0	93358_NCI-H292_IL-4	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	93360_NCI-H292_IL-9	0.0
93578_NK Cells IL-2_resting	0.0	93359_NCI-H292_IL-13	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93357_NCI-H292_IFN gamma	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93777_HPAEC_-	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93778_HPAEC_IL-1 beta/TNA alpha	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	93254_Normal Human Lung Fibroblast_none	34.1
93113_Mononuclear Cells (PBMCs)_PWM	0.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	93257_Normal Human Lung Fibroblast_IL-4	0.0
93249_Ramos (B cell)_none	0.0	93256_Normal Human Lung Fibroblast_IL-9	0.0
93250_Ramos (B cell)_ionomycin	0.0	93255_Normal Human Lung Fibroblast_IL-13	0.0
93349_B lymphocytes_PWM	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	13.0
93350_B lymphoytes_CD40L and IL-4	0.0	93106_Dermal Fibroblasts CCD1070_resting	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0
93356_Dendritic Cells_none	0.0	93772_dermal fibroblast_IFN gamma	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	93771_dermal fibroblast_IL-4	0.0
93775_Dendritic Cells_anti-CD40	45.9	93259_IBD Colitis 1**	10.0
93774_Monocytes_resting	0.0	93260_IBD Colitis 2	12.4
93776_Monocytes_LPS 50 ng/ml	0.0	93261_IBD Crohns	16.6
93581_Macrophages_resting	35.6	735010_Colon_normal	**100.0**
93582_Macrophages_LPS 100 ng/ml	0.0	735019_Lung_none	43.0
93098_HUVEC (Endothelial)_none	0.0	64028-1_Thymus_none	0.0
93099_HUVEC (Endothelial)_starved	0.0	64030-1_Kidney_none	0.0

Panel 1.3D Summary: <u>Ag1635</u> The expression of GPCR2 gene AP001804_B is low/undetectable (CT values >35) in all the tissues on this panel. <u>Ag2355</u> The expression of

the AP001804_B gene is low but significant in two breast cancer cell lines. Interestingly, the

two positive breast cancer cell lines are estrogen receptor positive. Thus, expression of this

gene may be indicative of estrogen receptor status on breast cancer cells and may have

implications to breast cancer cell biology. In addition, therapeutic modulation of this gene

5   may have utility in the treatment of breast cancer or other breast disease.

Panel 2 Summary: Ag2355 Expression of this gene is highest in a sample derived

from an ovarian cancer. Samples in which there is also expression are many fold lower than

the ovarian cancer. Thus, this gene may be useful for the diagnosis or therapeutic intervention

for ovarian cancer.

10   **Panel 2.2 Summary:** Ag1635 Expression of gene AP001804_B on this panel is too

low to be reliable (Ct values >35).

Panel 3D Summary: Ag2355 The expression of the AP001804_B gene in panel 3D

appears to be associated with lung cancer cell lines. Furthermore, the cell line that expresses

this gene in most abundance is neuroendocrine in origin. Neuroendocrine tumors are very

15   unique and thus, the AP001804_B gene may represent a unique marker of this type of cancer.

In addition, therapeutic modulation of this gene may be useful for the treatment of

neuroendocrine tumors in the lung.

Panel 4D Summary: Ag1635 The AP001804_B transcript is expressed in normal

colon but not in colons from patients with Crohn's disease or colitis. Protein therapeutics

20   designed with the putative GPCR encoded for by this gene could be used to inhibit

inflammation and tissue destruction due to IBD.

C. GPCR3 (also known as AP001804_C or CG54344-01)

Expression of gene AP001804_C was assessed using the primer-probe set Ag1639,

25   described in Table 20. Results of the RTQ-PCR runs are shown in Tables 21, 22, and 23.

Table 20. Probe Name Ag1639

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-AGCATCTTCCACATTGATTCC-3'	59	21	670	86
Probe	TET-5'-CTTCAGCACCTGCAGCTCCCACATAA-3'-TAMRA	71.2	26	711	87
Reverse	5'-CCAAAGAACAGAGAAACTGCAA-3'	59.5	22	737	88

Table 21. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5590t_ag1639_a2	Tissue Name	Relative Expression(%) 1.3dx4tm5590t_ag1639_a2
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	17.2
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	6.7	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.0	Mammary gland	0.0
CNS ca. (glio) U251	10.6	Breast ca.* (pl. effusion) MCF-7	15.2
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	25.7
Heart	0.0	Breast ca. BT-549	0.0
Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	**100.0**	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	6.6
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0

Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	6.1
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	37.4
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table 22. Panel 2.2

Tissue Name	Relative Expression(%) 2.2x4tm6361t_ ag1639_a2	Tissue Name	Relative Expression(%) 2.2x4tm6361t_ ag1639_a2
Normal Colon GENPAK 061003	17.6	83793 Kidney NAT (OD04348)	0.0
97759 Colon cancer (OD06064)	0.0	98938 Kidney malignant cancer (OD06204B)	0.0
97760 Colon cancer NAT (OD06064)	0.0	98939 Kidney normal adjacent tissue (OD06204E)	0.0
97778 Colon cancer (OD06159)	0.0	85973 Kidney Cancer (OD04450-01)	0.0
97779 Colon cancer NAT (OD06159)	0.0	85974 Kidney NAT (OD04450-03)	0.0
98861 Colon cancer (OD06297-04)	0.0	Kidney Cancer Clontech 8120613	0.0
98862 Colon cancer NAT (OD06297-015)	0.0	Kidney NAT Clontech 8120614	0.0
83237 CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer Clontech 9010320	0.0
83238 CC NAT (ODO3921)	0.0	Kidney NAT Clontech 9010321	0.0
97766 Colon cancer metastasis (OD06104)	0.0	Kidney Cancer Clontech 8120607	0.0
97767 Lung NAT (OD06104)	0.0	Kidney NAT Clontech 8120608	0.0
87472 Colon mets to lung (OD04451-01)	0.0	Normal Uterus GENPAK 061018	16.6
87473 Lung NAT (OD04451-02)	0.0	Uterus Cancer GENPAK 064011	0.0
Normal Prostate Clontech A+ 6546-1 (8090438)	0.0	Normal Thyroid Clontech A+ 6570-1 (7080817)	0.0
84140 Prostate Cancer (OD04410)	0.0	Thyroid Cancer GENPAK 064010	0.0
84141 Prostate NAT (OD04410)	0.0	Thyroid Cancer INVITROGEN A302152	0.0
Normal Ovary Res. Gen.	0.0	Thyroid NAT INVITROGEN A302153	0.0

98863 Ovarian cancer (OD06283-03)	0.0
98865 Ovarian cancer NAT/fallopian tube (OD06283-07)	0.0
Ovarian Cancer GENPAK 064008	**100.0**
97773 Ovarian cancer (OD06145)	0.0
97775 Ovarian cancer NAT (OD06145)	0.0
98853 Ovarian cancer (OD06455-03)	0.0
98854 Ovarian NAT (OD06455-07) Fallopian tube	0.0
Normal Lung GENPAK 061010	0.0
92337 Invasive poor diff. lung adeno (ODO4945-01	17.1
92338 Lung NAT (ODO4945-03)	0.0
84136 Lung Malignant Cancer (OD03126)	0.0
84137 Lung NAT (OD03126)	0.0
90372 Lung Cancer (OD05014A)	0.0
90373 Lung NAT (OD05014B)	0.0
97761 Lung cancer (OD06081)	0.0
97762 Lung cancer NAT (OD06081)	0.0
85950 Lung Cancer (OD04237-01)	0.0
85970 Lung NAT (OD04237-02)	0.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0
83256 Liver NAT (ODO4310)	0.0
84139 Melanoma Mets to Lung (OD04321)	0.0
84138 Lung NAT (OD04321)	0.0
Normal Kidney GENPAK 061008	0.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.0
83787 Kidney NAT (OD04338)	0.0

Normal Breast GENPAK 061019	0.0
84877 Breast Cancer (OD04566)	0.0
Breast Cancer Res. Gen. 1024	0.0
85975 Breast Cancer (OD04590-01)	0.0
85976 Breast Cancer Mets (OD04590-03)	0.0
87070 Breast Cancer Metastasis (OD04655-05)	0.0
GENPAK Breast Cancer 064006	0.0
Breast Cancer Clontech 9100266	0.0
Breast NAT Clontech 9100265	0.0
Breast Cancer INVITROGEN A209073	0.0
Breast NAT INVITROGEN A2090734	0.0
97763 Breast cancer (OD06083)	33.2
97764 Breast cancer node metastasis (OD06083)	0.0
Normal Liver GENPAK 061009	0.0
Liver Cancer Research Genetics RNA 1026	0.0
Liver Cancer Research Genetics RNA 1025	0.0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0
Paired Liver Tissue Research Genetics RNA 6004-N	0.0
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0
Paired Liver Tissue Research Genetics RNA 6005-N	0.0
Liver Cancer GENPAK 064003	72.5
Normal Bladder GENPAK 061001	0.0
Bladder Cancer Research Genetics RNA 1023	0.0
Bladder Cancer INVITROGEN A302173	0.0
Normal Stomach GENPAK 061017	0.0

83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer Clontech 9060397	0.0
83789 Kidney NAT (OD04339)	0.0	NAT Stomach Clontech 9060396	0.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer Clontech 9060395	17.0
83791 Kidney NAT (OD04340)	0.0	NAT Stomach Clontech 9060394	0.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer GENPAK 064005	0.0

Table 23. Panel 4D

Tissue Name	Relative Expression(%) 4dx4tm5519t_ ag1639_b2	Tissue Name	Relative Expression(%) 4dx4tm5519t_ ag1639_b2
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.4	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	93781_HUVEC (Endothelial)_IL-11	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	93583_Lung Microvascular Endothelial Cells_none	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	93347_Small Airway Epithelium_none	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	92668_Coronery Artery SMC_resting	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-	0.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b	0.0

160

CD3

		(1 ng/ml)	
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	93107_astrocytes_resting	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	92666_KU-812 (Basophil)_resting	0.0
93354_CD4_none	0.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.7
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	93579_CCD1106 (Keratinocytes)_none	0.0
		93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0
93103_LAK cells_resting	0.0	93791_Liver Cirrhosis	9.1
93788_LAK cells_IL-2	0.0	93792_Lupus Kidney	0.3
93787_LAK cells_IL-2+IL-12	0.0	93577_NCI-H292	0.0
93789_LAK cells_IL-2+IFN gamma	0.0	93358_NCI-H292_IL-4	0.0
93790_LAK cells_IL-2+ IL-18	0.0		
93104_LAK cells_PMA/ionomycin and IL-18	0.0	93360_NCI-H292_IL-9	0.0
93578_NK Cells IL-2_resting	0.0	93359_NCI-H292_IL-13	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93357_NCI-H292_IFN gamma	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93777_HPAEC_-	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93778_HPAEC_IL-1 beta/TNA alpha	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	93254_Normal Human Lung Fibroblast_none	0.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	93257_Normal Human Lung Fibroblast_IL-4	0.3
93249_Ramos (B cell)_none	0.0	93256_Normal Human Lung Fibroblast_IL-9	0.0
93250_Ramos (B cell)_ionomycin	0.0	93255_Normal Human Lung Fibroblast_IL-13	0.0
93349_B lymphocytes_PWM	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0
93350_B lymphoytes_CD40L and IL-4	0.0	93106_Dermal Fibroblasts CCD1070_resting	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0

		93772_dermal fibroblast_IFN	
93356_Dendritic Cells_none	1.1	gamma	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	93771_dermal fibroblast_IL-4	0.0
93775_Dendritic Cells_anti-CD40	1.7	93259_IBD Colitis 1**	100.0
93774_Monocytes_resting	0.0	93260_IBD Colitis 2	2.4
93776_Monocytes_LPS 50 ng/ml	0.0	93261_IBD Crohns	0.0
93581_Macrophages_resting	0.4	735010_Colon_normal	5.6
93582_Macrophages_LPS 100 ng/ml	0.0	735019_Lung_none	0.0
93098_HUVEC (Endothelial)_none	0.0	64028-1_Thymus_none	0.0
93099_HUVEC (Endothelial)_starved	0.0	64030-1_Kidney_none	0.0

Panel 1.3D Summary: Expression of the AP001804_C gene in this panel is highest in the spleen. Expression is detected at a much lower level in a melanoma and a breast cancer cell line. This profile may indicate that the expression of this gene is restricted to splenic lymphoid tissues and thus may be useful as a marker of this tissue.

Panel 2.2 Summary: Significant expression of the AP001804_C gene on panel 2.2 is restricted to one ovarian cancer and one liver cancer. This information suggests that this gene may be of use in the diagnosis and/or treatment of ovarian or liver cancer.

Panel 4D Summary: The AP001804_C transcript is expressed in colitis 1, colitis 2, an activated basophil cell line and in dendritic cells. The protein encoded for by this antigen may be important in the inflammatory process and particularly in the function of activated dendritic cells or basophils. Antagonistic antibodies or small molecule therapeutics against the AP001804_C protein may therefore reduce or inhibit inflammation in the bowel due to IBD by specifically targeting dendritic cells and basophils or other related cell types. This gene was found to be expressed in spleen in Panel 1.3D.

D. GPCR4 (also known as AP001804_D or CG54353-01)

Expression of gene AP001804_D was assessed using the primer-probe set Ag3091, described in Table 24. Results of the RTQ-PCR runs are shown in Tables 25, 26, and 27.

Table 24. Probe Name Ag3091

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GGTGCATGACTCAGCTGTTT-3'	58.9	20	287	89

					90
Probe	FAM-5'-TCATCTCTGAATGTTACATGTTGACCTCA-3'-TAMRA	65.9	29	323	
Reverse	5'-GCCACATAGCGATCATATGC-3'	59.1	20	355	91

Table 25. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5823f _ag3091_b1	Tissue Name	Relative Expression(%) 1.3dx4tm5823f _ag3091_b1
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	1.6
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	8.1	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	2.2
CNS ca. (glio/astro) U-118-MG	4.7	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.0	Mammary gland	0.0
CNS ca. (glio) U251	0.0	Breast ca.* (pl. effusion) MCF-7	74.9
CNS ca. (glio) SF-295	1.2	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	100.0
Heart	0.0	Breast ca. BT-549	0.0
Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	15.7
Spleen	0.0	Ovarian ca. OVCAR-5	0.0

163

Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	5.5	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	8.3
Colon ca. HCT-116	0.0	Testis	16.3
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	20.5
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table 26. Panel 2.2

Tissue Name	Relative Expression(%) 2.2x4tm6415f_ ag3091_b1	Tissue Name	Relative Expression(%) 2.2x4tm6415f_ ag3091_b1
Normal Colon GENPAK 061003	0.0	83793 Kidney NAT (OD04348)	0.0
97759 Colon cancer (OD06064)	0.0	98938 Kidney malignant cancer (OD06204B)	0.0
97760 Colon cancer NAT (OD06064)	0.0	98939 Kidney normal adjacent tissue (OD06204E)	10.7
97778 Colon cancer (OD06159)	0.0	85973 Kidney Cancer (OD04450-01)	0.0
97779 Colon cancer NAT (OD06159)	0.0	85974 Kidney NAT (OD04450-03)	0.0
98861 Colon cancer (OD06297-04)	0.0	Kidney Cancer Clontech 8120613	0.0
98862 Colon cancer NAT (OD06297-015)	0.0	Kidney NAT Clontech 8120614	0.0
83237 CC Gr.2 ascend colon (ODO3921)	0.0	Kidney Cancer Clontech 9010320	0.0
83238 CC NAT (ODO3921)	0.0	Kidney NAT Clontech 9010321	15.6
97766 Colon cancer metastasis (OD06104)	0.0	Kidney Cancer Clontech 8120607	0.0
97767 Lung NAT (OD06104)	8.7	Kidney NAT Clontech 8120608	0.0
87472 Colon mets to lung (OD04451-01)	0.0	Normal Uterus GENPAK 061018	13.7
87473 Lung NAT (OD04451-02)	0.0	Uterus Cancer GENPAK 064011	0.0
Normal Prostate Clontech A+ 6546-1 (8090438)	0.0	Normal Thyroid Clontech A+ 6570-1 (7080817)	0.0

84140 Prostate Cancer (OD04410)	0.0	Thyroid Cancer GENPAK 064010	0.0
84141 Prostate NAT (OD04410)	0.0	Thyroid Cancer INVITROGEN A302152	0.0
		Thyroid NAT INVITROGEN A302153	0.0
Normal Ovary Res. Gen.	0.0	Normal Breast GENPAK 061019	0.0
98863 Ovarian cancer (OD06283-03)	0.0		
98865 Ovarian cancer NAT/fallopian tube (OD06283-07)	0.0	84877 Breast Cancer (OD04566)	0.0
Ovarian Cancer GENPAK 064008	4.1	Breast Cancer Res. Gen. 1024	0.0
97773 Ovarian cancer (OD06145)	0.0	85975 Breast Cancer (OD04590-01)	0.0
97775 Ovarian cancer NAT (OD06145)	22.8	85976 Breast Cancer Mets (OD04590-03)	0.0
98853 Ovarian cancer (OD06455-03)	0.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0
98854 Ovarian NAT (OD06455-07) Fallopian tube	0.0	GENPAK Breast Cancer 064006	0.0
		Breast Cancer Clontech 9100266	0.0
Normal Lung GENPAK 061010	0.0		
92337 Invasive poor diff. lung adeno (ODO4945-01	11.2	Breast NAT Clontech 9100265	0.0
92338 Lung NAT (ODO4945-03)	0.0	Breast Cancer INVITROGEN A209073	0.0
84136 Lung Malignant Cancer (OD03126)	0.0	Breast NAT INVITROGEN A2090734	0.0
84137 Lung NAT (OD03126)	0.0	97763 Breast cancer (OD06083)	0.0
90372 Lung Cancer (OD05014A)	0.0	97764 Breast cancer node metastasis (OD06083)	0.0
90373 Lung NAT (OD05014B)	4.0	Normal Liver GENPAK 061009	0.0
97761 Lung cancer (OD06081)	0.0	Liver Cancer Research Genetics RNA 1026	0.0
97762 Lung cancer NAT (OD06081)	0.0	Liver Cancer Research Genetics RNA 1025	0.0
85950 Lung Cancer (OD04237-01)	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0
85970 Lung NAT (OD04237-02)	0.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0
83256 Liver NAT (ODO4310)	0.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0
84139 Melanoma Mets to Lung (OD04321)	0.0	Liver Cancer GENPAK 064003	14.5
84138 Lung NAT (OD04321)	0.0	Normal Bladder GENPAK 061001	0.0

Normal Kidney GENPAK 061008	0.0	Bladder Cancer Research Genetics RNA 1023	0.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	13.0	Bladder Cancer INVITROGEN A302173	**100.0**
83787 Kidney NAT (OD04338)	14.8	Normal Stomach GENPAK 061017	0.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	Gastric Cancer Clontech 9060397	0.0
83789 Kidney NAT (OD04339)	0.0	NAT Stomach Clontech 9060396	5.4
83790 Kidney Ca, Clear cell type (OD04340)	0.0	Gastric Cancer Clontech 9060395	15.0
83791 Kidney NAT (OD04340)	0.0	NAT Stomach Clontech 9060394	0.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	Gastric Cancer GENPAK 064005	0.0

<u>Table 27</u>. Panel 4D

Tissue Name	Relative Expression(%) 4dx4tm5055f_ ag3091_b2	Tissue Name	Relative Expression(%) 4dx4tm5055f_ ag3091_b2
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	93781_HUVEC (Endothelial)_IL-11	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	93583_Lung Microvascular Endothelial Cells_none	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	93347_Small Airway Epithelium_none	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93351_CD45RA CD4	0.0	92668_Coronery Artery	0.0

lymphocyte_anti-CD28/anti-CD3		SMC_resting	
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	93107_astrocytes_resting	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	92666_KU-812 (Basophil)_resting	0.0
93354_CD4_none	0.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	93579_CCD1106 (Keratinocytes)_none	0.0
		93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0
93103_LAK cells_resting	0.0	93791_Liver Cirrhosis	0.0
93788_LAK cells_IL-2	0.0	93792_Lupus Kidney	0.0
93787_LAK cells_IL-2+IL-12	0.0		
93789_LAK cells_IL-2+IFN gamma	0.0	93577_NCI-H292	0.0
93790_LAK cells_IL-2+ IL-18	0.0	93358_NCI-H292_IL-4	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	93360_NCI-H292_IL-9	0.0
93578_NK Cells IL-2_resting	0.0	93359_NCI-H292_IL-13	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93357_NCI-H292_IFN gamma	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93777_HPAEC_-	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	93778_HPAEC_IL-1 beta/TNA alpha	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	93254_Normal Human Lung Fibroblast_none	0.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	93257_Normal Human Lung Fibroblast_IL-4	17.0
93249_Ramos (B cell)_none	0.0	93256_Normal Human Lung Fibroblast_IL-9	0.0
93250_Ramos (B cell)_ionomycin	0.0	93255_Normal Human Lung Fibroblast_IL-13	0.0
93349_B lymphocytes_PWM	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0
93350_B lymphoytes_CD40L and IL-4	0.0	93106_Dermal Fibroblasts CCD1070_resting	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0

167

93248_EOL-1 (Eosinophil)_dbcAMP/PMAion omycin	0.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0
93356_Dendritic Cells_none	31.7	93772_dermal fibroblast_IFN gamma	0.0
93355_Dendritic Cells_LPS 100 ng/ml	4.6	93771_dermal fibroblast_IL-4	0.0
93775_Dendritic Cells_anti-CD40	43.3	93259_IBD Colitis 1**	0.0
93774_Monocytes_resting	0.0	93260_IBD Colitis 2	2.9
93776_Monocytes_LPS 50 ng/ml	0.0	93261_IBD Crohns	7.7
93581_Macrophages_resting	100.0	735010_Colon_normal	40.3
93582_Macrophages_LPS 100 ng/ml	0.0	735019_Lung_none	55.9
93098_HUVEC (Endothelial)_none	0.0	64028-1_Thymus_none	0.0
93099_HUVEC (Endothelial)_starved	0.0	64030-1_Kidney_none	0.0

Panel 1.3D Summary: The expression of the AP001804_D gene appears to be restricted to two breast cancer cell lines. Interestingly both of these cell lines are positive for estrogen receptor expression. Thus, this gene may be a marker for estrogen receptor positive breast cancer cells. Further, therapeutic modulation of this gene may be of use in the treatment of breast cancer or other breast related disease.

Panel 2.2 Summary: Two RTQ-PCR experiments were performed using Ag3091. In one experiment, AP001804_D gene expression was low to undetectable (CT values >35) in all samples. In the other experiment, expression was low/undectable in all samples except a single bladder cancer cell line (CT=34.5). Expression levels are too low for reliable analysis.

Panel 4D Summary: The AP001804_D transcript is detectable in resting macrophages and not at significant levels in other cell types. Antibody or protein therapeutics designed against the AP001804_D protein encoded for by this transcript could reduce or inhibit inflammation in asthma, emphysema, allergy, psoriasis, arthritis, or any other condition in which macrophage localization/activation is important.

E. GPCR5 (also known as AP001804_E or CG54362-01)

Expression of gene AP001804_E was assessed using the primer-probe sets Ag2359, Ag2358, and Ag1640 (identical sequences), described in Table 28. Results of the RTQ-PCR runs are shown in Tables 29 and 30.

Table 28. Probe Name Ag2359/Ag2358/Ag1640

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CCATGTCAGTGAGCTGGTATTT-3'	59.1	22	574	92
Probe	FAM-5'-TGGAGTAATCACCATGCTATCCAGCA-3'-TAMRA	67.7	26	607	93
Reverse	5'-TCAAAGCGTAAGAGATGACGAT-3'	59	22	638	94

Table 29. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5396f_ag1640_a2	Tissue Name	Relative Expression(%) 1.3dx4tm5396f_ag1640_a2
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	39.2
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	0.0	Liver (fetal)	0.0
Brain (amygdala)	0.0	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	0.0	Lung	0.0
Brain (hippocampus)	0.0	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (thalamus)	38.8	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.0	Mammary gland	0.0
CNS ca. (glio) U251	20.2	Breast ca.* (pl. effusion) MCF-7	87.4
CNS ca. (glio) SF-295	21.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	76.7

Heart	21.3	Breast ca. BT-549	0.0
Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	0.0	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	0.0	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	44.8
Colon ca. HCT-116	0.0	Testis	16.4
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	**100.0**
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0
Trachea	0.0	Melanoma* (met) SK-MEL-5	0.0
Kidney	0.0	Adipose	0.0

Table 30. Panel 2D

Tissue Name	Relative Expression(%) 2dx4tm4937f_ag2359_a1	Relative Expression(%) 2dx4tm4923f_ag2358_b2
Normal Colon GENPAK 061003	2.3	8.0
83219 CC Well to Mod Diff (ODO3866)	4.2	4.4
83220 CC NAT (ODO3866)	9.2	1.7
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	0.0
83222 CC NAT (ODO3868)	0.0	0.0
83235 CC Mod Diff (ODO3920)	0.0	0.0
83236 CC NAT (ODO3920)	0.0	0.0
83237 CC Gr.2 ascend colon (ODO3921)	0.0	0.0
83238 CC NAT (ODO3921)	1.2	1.1
83241 CC from Partial Hepatectomy (ODO4309)	0.0	0.0
83242 Liver NAT (ODO4309)	0.0	0.0
87472 Colon mets to lung (OD04451-01)	0.0	4.5
87473 Lung NAT (OD04451-02)	0.0	0.0
Normal Prostate Clontech A+ 6546-1	0.0	0.0
84140 Prostate Cancer (OD04410)	0.0	0.0
84141 Prostate NAT (OD04410)	0.0	0.0

87073 Prostate Cancer (OD04720-01)	0.0	0.0
87074 Prostate NAT (OD04720-02)	0.0	0.0
Normal Lung GENPAK 061010	0.0	0.0
83239 Lung Met to Muscle (OD04286)	2.1	0.0
83240 Muscle NAT (OD04286)	0.0	0.0
84136 Lung Malignant Cancer (OD03126)	0.0	1.3
84137 Lung NAT (OD03126)	0.0	0.9
84871 Lung Cancer (OD04404)	0.0	0.0
84872 Lung NAT (OD04404)	0.0	0.0
84875 Lung Cancer (OD04565)	0.0	0.0
84876 Lung NAT (OD04565)	0.0	0.0
85950 Lung Cancer (OD04237-01)	0.0	1.7
85970 Lung NAT (OD04237-02)	0.0	0.0
83255 Ocular Mel Met to Liver (OD04310)	1.8	0.0
83256 Liver NAT (OD04310)	0.0	0.0
84139 Melanoma Mets to Lung (OD04321)	0.0	0.0
84138 Lung NAT (OD04321)	0.0	0.0
Normal Kidney GENPAK 061008	3.4	5.1
83786 Kidney Ca, Nuclear grade 2 (OD04338)	8.1	13.4
83787 Kidney NAT (OD04338)	0.0	4.1
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	5.6
83789 Kidney NAT (OD04339)	1.6	0.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	0.0
83791 Kidney NAT (OD04340)	3.7	0.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	0.0
83793 Kidney NAT (OD04348)	0.0	0.0
87474 Kidney Cancer (OD04622-01)	0.0	0.0
87475 Kidney NAT (OD04622-03)	0.0	0.0
85973 Kidney Cancer (OD04450-01)	0.0	0.0
85974 Kidney NAT (OD04450-03)	0.0	0.0
Kidney Cancer Clontech 8120607	0.0	0.0
Kidney NAT Clontech 8120608	0.0	0.0
Kidney Cancer Clontech 8120613	0.0	0.0
Kidney NAT Clontech 8120614	0.0	0.0
Kidney Cancer Clontech 9010320	0.0	0.0
Kidney NAT Clontech 9010321	0.0	0.0
Normal Uterus GENPAK 061018	0.0	0.8
Uterus Cancer GENPAK 064011	0.0	0.0
Normal Thyroid Clontech A+ 6570-1	0.0	0.0
Thyroid Cancer GENPAK 064010	0.0	0.0
Thyroid Cancer INVITROGEN A302152	0.0	0.0
Thyroid NAT INVITROGEN A302153	0.0	0.0
Normal Breast GENPAK 061019	0.0	0.0
84877 Breast Cancer (OD04566)	0.0	0.0
85975 Breast Cancer (OD04590-01)	0.0	0.0
85976 Breast Cancer Mets (OD04590-03)	2.2	0.0

87070 Breast Cancer Metastasis (OD04655-05)	0.0	0.0
GENPAK Breast Cancer 064006	1.7	0.0
Breast Cancer Res. Gen. 1024	0.0	3.3
Breast Cancer Clontech 9100266	0.0	0.0
Breast NAT Clontech 9100265	0.0	0.0
Breast Cancer INVITROGEN A209073	0.0	1.8
Breast NAT INVITROGEN A2090734	0.0	0.0
Normal Liver GENPAK 061009	0.0	0.0
Liver Cancer GENPAK 064003	21.7	6.5
Liver Cancer Research Genetics RNA 1025	0.0	0.0
Liver Cancer Research Genetics RNA 1026	0.0	0.0
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	1.4	0.0
Paired Liver Tissue Research Genetics RNA 6004-N	0.0	0.0
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	0.0
Paired Liver Tissue Research Genetics RNA 6005-N	0.0	0.0
Normal Bladder GENPAK 061001	0.0	0.0
Bladder Cancer Research Genetics RNA 1023	0.0	0.0
Bladder Cancer INVITROGEN A302173	45.9	57.8
87071 Bladder Cancer (OD04718-01)	0.0	0.0
87072 Bladder Normal Adjacent (OD04718-03)	0.0	0.0
Normal Ovary Res. Gen.	0.0	0.0
Ovarian Cancer GENPAK 064008	0.0	0.0
87492 Ovary Cancer (OD04768-07)	**100.0**	**100.0**
87493 Ovary NAT (OD04768-08)	0.0	0.0
Normal Stomach GENPAK 061017	3.5	0.0
Gastric Cancer Clontech 9060358	0.0	0.0
NAT Stomach Clontech 9060359	0.0	0.0
Gastric Cancer Clontech 9060395	0.0	0.0
NAT Stomach Clontech 9060394	0.0	0.0
Gastric Cancer Clontech 9060397	0.0	0.0
NAT Stomach Clontech 9060396	0.0	0.0
Gastric Cancer GENPAK 064005	0.0	1.8

Panel 1.3 D Summary: Ag1640 Significant expression of the AP001804_E gene is restricted to one melanoma cell line indicating that this gene may be a useful marker for melanoma. Ag2358/Ag2359 Expression of the AP001804_E gene was low/undetectable (CT values > 35) in all samples on this panel.

Panel 2D Summary: Ag2359/Ag2358 The AP001804_E gene is most abundantly expressed in a sample of ovarian cancer with limited, very low level of expression in other tissues. Thus, this gene may be useful in distinguishing ovarian cancers from other tissues. Therapeutic modulation of this gene may also be useful in the treatment of ovarian cancers.

172

Panel 2.2 Summary: <u>Ag1640</u> Expression of the AP001804_E gene was

low/undetectable (CT values > 35) in all samples on this panel and thus the results not shown.

Panel 4D Summary: <u>Ag2359/Ag2358/Ag1640</u> Expression of the AP001804_E gene

was low/undetectable (CT values > 35) in all samples on this panel and thus the results not

shown.

F. GPCR6 (also known as AP000868_A or CG54263-01)

Expression of gene AP000868_A was assessed using the primer-probe set Ag1629,

described in Table 31. Results of the RTQ-PCR run is shown in Table 32.

<u>Table 31</u>. **Probe Name Ag1629**

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-AAATCTGGTACACCACCACAGT-3'	58.4	22	203	95
Probe	FAM-5'-CATCCCCAAACTGCTAGGAACCTTTG-3'-TAMRA	68.5	26	225	96
Reverse	5'-AGCAGGACATGCAGATTACTGT-3'	58.9	22	262	97

<u>Table 32</u>. **Panel 1.3D**

Tissue Name	Relative Expression(%) 1.3dx4tm5395f _ag1629_b2	Tissue Name	Relative Expression(%) 1.3dx4tm5395f _ag1629_b2
Liver adenocarcinoma	0.0	Kidney (fetal)	0.0
Pancreas	21.2	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	0.0
Adrenal gland	0.0	Renal ca. RXF 393	0.0
Thyroid	0.0	Renal ca. ACHN	0.0
Salivary gland	0.0	Renal ca. UO-31	0.0
Pituitary gland	0.0	Renal ca. TK-10	0.0
Brain (fetal)	0.0	Liver	0.0
Brain (whole)	5.7	Liver (fetal)	0.0
Brain (amygdala)	78.2	Liver ca. (hepatoblast) HepG2	0.0
Brain (cerebellum)	22.7	Lung	0.0
Brain (hippocampus)	**100.0**	Lung (fetal)	0.0
Brain (substantia nigra)	0.0	Lung ca. (small cell) LX-1	0.0

173

Brain (thalamus)	0.0	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	0.0	Lung ca. (large cell)NCI-H460	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca. (non-s.cell) NCI-H23	4.1
CNS ca. (astro) SW1783	0.0	Lung ca (non-s.cell) HOP-62	0.0
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SNB-75	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.0	Mammary gland	0.0
CNS ca. (glio) U251	11.1	Breast ca.* (pl. effusion) MCF-7	0.0
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	0.0	Breast ca.* (pl. effusion) T47D	0.0
Heart	0.0	Breast ca. BT-549	0.0
Fetal Skeletal	0.0	Breast ca. MDA-N	0.0
Skeletal muscle	0.0	Ovary	0.0
Bone marrow	0.0	Ovarian ca. OVCAR-3	0.0
Thymus	0.0	Ovarian ca. OVCAR-4	0.0
Spleen	0.0	Ovarian ca. OVCAR-5	0.0
Lymph node	3.3	Ovarian ca. OVCAR-8	0.0
Colorectal	0.0	Ovarian ca. IGROV-1	0.0
Stomach	2.5	Ovarian ca.* (ascites) SK-OV-3	0.0
Small intestine	0.0	Uterus	0.0
Colon ca. SW480	0.0	Placenta	0.0
Colon ca.* (SW480 met)SW620	0.0	Prostate	0.0
Colon ca. HT29	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	0.0
Colon ca. CaCo-2	0.0	Melanoma Hs688(A).T	0.0
83219 CC Well to Mod Diff (ODO3866)	0.0	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.0	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma M14	0.0
Bladder	0.0	Melanoma LOX IMVI	0.0

| Trachea | 0.0 | Melanoma* (met) SK-MEL-5 | 0.0 |
| Kidney | 0.0 | Adipose | 9.1 |

Panel 1.3D Summary: The AP000868_A transcript appears to be brain specific (or at least to show highly preferential expression in brain), especially in the hippocampus and amygdala. These regions are of great interest as both have been implicated in Alzheimer's disease, schizophrenia, and bipolar disorder. Furthermore, the hippocampus is critical in the development of long-term memories, and the amygdala is involved in the processing of emotion (e.g., fear, etc). Because this transcript encodes for a GPCR, the AP000868_A protein is also a potential small molecule target for the treatment/prevention of both neurodegenerative and psychiatric disorders. In addition, the AP000868_A gene product could possibly be targeted in normal, healthy populations for modulation of memory and fear/anxiety.

Panel 2.2 Summary: Expression of the AP000868_A gene was low/undetectable (CT values > 35) in all samples on this panel and thus the results not shown.

Panel 4D Summary: Expression of the AP000868_A gene was low/undetectable (CT values > 35) in all samples on this panel and thus the results not shown.

Panel CNSD.01 Summary: Expression of the AP000868_A gene was low/undetectable (CT values > 35) in all samples on this panel and thus the results not shown.

G. GPCR7 (also known as 20722608_EXT or CG51505-01)

Expression of gene 20722608_EXT was assessed using the primer-probe set Ag1629, described in Table 33. Results of the RTQ-PCR run is shown in Table 34.

<u>Table 33</u>. Probe Name Gpcr27

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TGCCCTTGTCATCCTGAGC-3'		19	187	98
Probe	FAM-5'-CCTGGCATTCACTACTGGGTCCTACGTGTA-3'-TAMRA		30	208	99
Reverse	5'-GGGATACGCAGGATGGTAGAAA-3'		22	243	100

Table 34. Panel 1

Tissue Name	Relative Expression(%) 1xtm240f_ gpcr27	Tissue Name	Relative Expression(%) 1xtm240f_ gpcr27
Endothelial cells	2.6	Kidney (fetal)	2.5
Endothelial cells (treated)	5.2	Renal ca. 786-0	0
Pancreas	100	Renal ca. A498	0.6
Pancreatic ca. CAPAN 2	0	Renal ca. RXF 393	0.7
Adipose	10.4	Renal ca. ACHN	0
Adrenal gland	0	Renal ca. UO-31	2.4
Thyroid	0.7	Renal ca. TK-10	0
Salavary gland	0	Liver	0
Pituitary gland	0	Liver (fetal)	0
Brain (fetal)	3.3	Liver ca. (hepatoblast) HepG2	0
Brain (whole)	5.4	Lung	0
Brain (amygdala)	2.2	Lung (fetal)	0
Brain (cerebellum)	1.2	Lung ca. (small cell) LX-1	0
Brain (hippocampus)	3	Lung ca. (small cell) NCI-H69	11.7
Brain (substantia nigra)	0.6	Lung ca. (s.cell var.) SHP-77	0
Brain (thalamus)	0	Lung ca. (large cell)NCI-H460	0
Brain (hypothalamus)	0.2	Lung ca. (non-sm. cell) A549	0
Spinal cord	0	Lung ca. (non-s.cell) NCI-H23	0
CNS ca. (glio/astro) U87-MG	1.5	Lung ca (non-s.cell) HOP-62	0.6
CNS ca. (glio/astro) U-118-MG	0	Lung ca. (non-s.cl) NCI-H522	0
CNS ca. (astro) SW1783	1.5	Lung ca. (squam.) SW 900	1.3
CNS ca.* (neuro; met) SK-N-AS	0	Lung ca. (squam.) NCI-H596	1.4
CNS ca. (astro) SF-539	0	Mammary gland	1.4
CNS ca. (astro) SNB-75	0	Breast ca.* (pl. effusion) MCF-7	0
CNS ca. (glio) SNB-19	1.2	Breast ca.* (pl.ef) MDA-MB-231	0
CNS ca. (glio) U251	0.6	Breast ca.* (pl. effusion) T47D	21.9
CNS ca. (glio) SF-295	0	Breast ca. BT-549	0
Heart	0	Breast ca. MDA-N	1.5
Skeletal muscle	0	Ovary	0
Bone marrow	0	Ovarian ca. OVCAR-3	0
Thymus	0.6	Ovarian ca. OVCAR-4	1.5
Spleen	1.5	Ovarian ca. OVCAR-5	5.6
Lymph node	0	Ovarian ca. OVCAR-8	0.7
Colon (ascending)	57	Ovarian ca. IGROV-1	2
Stomach	0	Ovarian ca.* (ascites) SK-OV-3	0
Small intestine	0	Uterus	2.2
Colon ca. SW480	0	Placenta	4.1
Colon ca.* (SW480 met)SW620	0	Prostate	0

Colon ca. HT29	0.8	Prostate ca.* (bone met)PC-3	0
Colon ca. HCT-116	0	Testis	99.3
Colon ca. CaCo-2	0.8	Melanoma Hs688(A).T	0
Colon ca. HCT-15	2.9	Melanoma* (met) Hs688(B).T	3.4
Colon ca. HCC-2998	0	Melanoma UACC-62	0
Gastric ca.* (liver met) NCI-N87	2.3	Melanoma M14	10.6
Bladder	0	Melanoma LOX IMVI	1.3
Trachea	1.2	Melanoma* (met) SK-MEL-5	0.2
Kidney	1	Melanoma SK-MEL-28	0.3

Panel 1 Summary: The 20722608_EXT gene is expressed most abundantly in testis, colon and pancreas. Expression in the testis may be due to genomic DNA contamination. The expression of 20722608_EXT gene seems to be specific for pancreas and colon tissues. These tissues both play an important role in the process of digestion and thus, therapeutic modulation of the 20722608_EXT gene may be of utility in the treatment of gastrointestinal disease related to the colon and/or pancreas. In addition, although the 20722608_EXT gene is most highly expressed in the pancreas, it is absent in a pancreatic cancer cell lines suggesting that this gene could be useful in the diagnosis/treatment of pancreatic cancer. 20722608_EXT gene may be involved in signal transduction pathways in either the exocrine or endocrine tissues of the pancreas. Thus, this gene may be a drug target for diseases of the pancreas including Types 1 and 2 diabetes and any or all forms of pancreatitis.

H. 21629632_EXT

Expression of gene 21629632_EXT was assessed using the primer-probe set Ag1539, described in Table 35. Results of the RTQ-PCR run is shown in Table 36, 37, 38, 39 and 40.

Table 35. Probe name Ag1539

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TTTTATGGGACAATCTCCTTCA-3'	58.6	22	745	101
Probe	FAM-5'-TGTACTTCAAACCCAAGGCCAAGGAT-3'-TAMRA	68.4	26	767	102
Reverse	5'-GAACAATGCGACAGTCTTATCC-3'	58.7	22	801	103

177

Table 36. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm2212f_ag1539	Tissue Name	Relative Expression(%) 1.2tm2212f_ag1539
Endothelial cells	0.1	Renal ca. 786-0	0.7
Endothelial cells (treated)	3.5	Renal ca. A498	3.1
Pancreas	2.7	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	3.4
Adrenal Gland (new lot*)	9.5	Renal ca. UO-31	2.2
Thyroid	1.3	Renal ca. TK-10	3.1
Salavary gland	26.1	Liver	2.8
Pituitary gland	2.1	Liver (fetal)	2.6
Brain (fetal)	4.9	Liver ca. (hepatoblast) HepG2	0.5
Brain (whole)	22.8	Lung	0.5
Brain (amygdala)	14.9	Lung (fetal)	0.8
Brain (cerebellum)	14.0	Lung ca. (small cell) LX-1	13.0
Brain (hippocampus)	81.2	Lung ca. (small cell) NCI-H69	2.0
Brain (thalamus)	31.9	Lung ca. (s.cell var.) SHP-77	0.1
Cerebral Cortex	**100.0**	Lung ca. (large cell)NCI-H460	2.8
Spinal cord	3.3	Lung ca. (non-sm. cell) A549	4.1
CNS ca. (glio/astro) U87-MG	1.4	Lung ca. (non-s.cell) NCI-H23	1.2
CNS ca. (glio/astro) U-118-MG	0.3	Lung ca (non-s.cell) HOP-62	8.4
CNS ca. (astro) SW1783	0.4	Lung ca. (non-s.cl) NCI-H522	23.3
CNS ca.* (neuro; met) SK-N-AS	1.7	Lung ca. (squam.) SW 900	13.8
CNS ca. (astro) SF-539	1.7	Lung ca. (squam.) NCI-H596	1.3
CNS ca. (astro) SNB-75	1.9	Mammary gland	6.6
CNS ca. (glio) SNB-19	5.0	Breast ca.* (pl. effusion) MCF-7	1.2
CNS ca. (glio) U251	3.1	Breast ca.* (pl.ef) MDA-MB-231	0.5
CNS ca. (glio) SF-295	25.9	Breast ca.* (pl. effusion) T47D	5.4
Heart	46.3	Breast ca. BT-549	37.4
Skeletal Muscle (new lot*)	52.1	Breast ca. MDA-N	1.3
Bone marrow	0.4	Ovary	7.1
Thymus	0.3	Ovarian ca. OVCAR-3	3.7
Spleen	1.2	Ovarian ca. OVCAR-4	1.8
Lymph node	0.6	Ovarian ca. OVCAR-5	27.7
Colorectal	0.2	Ovarian ca. OVCAR-8	6.6
Stomach	2.5	Ovarian ca. IGROV-1	5.7
Small intestine	7.1	Ovarian ca.* (ascites) SK-OV-3	3.4
Colon ca. SW480	0.3	Uterus	3.2
Colon ca.* (SW480 met)SW620	0.9	Placenta	0.4
Colon ca. HT29	1.5	Prostate	20.2
Colon ca. HCT-116	0.9	Prostate ca.* (bone met)PC-3	3.3
Colon ca. CaCo-2	2.3	Testis	1.3

83219 CC Well to Mod Diff (ODO3866)	0.6	Melanoma Hs688(A).T	0.6
Colon ca. HCC-2998	11.9	Melanoma* (met) Hs688(B).T	0.5
Gastric ca.* (liver met) NCI-N87	4.9	Melanoma UACC-62	3.9
Bladder	5.0	Melanoma M14	1.6
Trachea	0.2	Melanoma LOX IMVI	0.0
Kidney	30.4	Melanoma* (met) SK-MEL-5	0.0
Kidney (fetal)	11.5	Adipose	18.0

Table 37. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3Dtm2998f_ ag1539	Tissue Name	Relative Expression(%) 1.3Dtm2998f_ ag1539
Liver adenocarcinoma	1.7	Kidney (fetal)	1.8
Pancreas	0.5	Renal ca. 786-0	1.6
Pancreatic ca. CAPAN 2	0.0	Renal ca. A498	2.2
Adrenal gland	2.7	Renal ca. RXF 393	0.5
Thyroid	4.4	Renal ca. ACHN	1.7
Salivary gland	2.0	Renal ca. UO-31	0.0
Pituitary gland	7.4	Renal ca. TK-10	1.2
Brain (fetal)	21.6	Liver	0.2
Brain (whole)	26.6	Liver (fetal)	1.6
Brain (amygdala)	30.8	Liver ca. (hepatoblast) HepG2	0.9
Brain (cerebellum)	7.6	Lung	1.7
Brain (hippocampus)	**100.0**	Lung (fetal)	3.5
Brain (substantia nigra)	5.0	Lung ca. (small cell) LX-1	4.1
Brain (thalamus)	15.8	Lung ca. (small cell) NCI-H69	1.2
Cerebral Cortex	76.8	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	4.3	Lung ca. (large cell)NCI-H460	0.3
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	2.1
CNS ca. (glio/astro) U-118-MG	0.3	Lung ca. (non-s.cell) NCI-H23	0.6
CNS ca. (astro) SW1783	0.7	Lung ca (non-s.cell) HOP-62	2.2
CNS ca.* (neuro; met) SK-N-AS	0.9	Lung ca. (non-s.cl) NCI-H522	4.0
CNS ca. (astro) SF-539	1.5	Lung ca. (squam.) SW 900	2.6
CNS ca. (astro) SNB-75	3.6	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.7	Mammary gland	1.8
CNS ca. (glio) U251	3.6	Breast ca.* (pl. effusion) MCF-7	0.0
CNS ca. (glio) SF-295	15.6	Breast ca.* (pl.ef) MDA-MB-231	0.9
Heart (fetal)	6.1	Breast ca.* (pl. effusion) T47D	1.6
Heart	2.4	Breast ca. BT-549	0.8
Fetal Skeletal	70.7	Breast ca. MDA-N	0.0

179

Skeletal muscle	0.5	Ovary	7.7
Bone marrow	0.0	Ovarian ca. OVCAR-3	1.0
Thymus	1.1	Ovarian ca. OVCAR-4	0.0
Spleen	0.4	Ovarian ca. OVCAR-5	4.8
Lymph node	1.0	Ovarian ca. OVCAR-8	1.8
Colorectal	8.5	Ovarian ca. IGROV-1	1.1
Stomach	2.9	Ovarian ca.* (ascites) SK-OV-3	0.6
Small intestine	4.5	Uterus	4.0
Colon ca. SW480	0.0	Placenta	0.3
Colon ca.* (SW480 met)SW620	0.9	Prostate	4.7
Colon ca. HT29	1.1	Prostate ca.* (bone met)PC-3	2.4
Colon ca. HCT-116	0.1	Testis	5.0
Colon ca. CaCo-2	0.9	Melanoma Hs688(A).T	1.3
83219 CC Well to Mod Diff (ODO3866)	1.2	Melanoma* (met) Hs688(B).T	1.8
Colon ca. HCC-2998	1.8	Melanoma UACC-62	0.7
Gastric ca.* (liver met) NCI-N87	3.3	Melanoma M14	0.3
Bladder	4.2	Melanoma LOX IMVI	0.0
Trachea	2.3	Melanoma* (met) SK-MEL-5	0.4
Kidney	3.3	Adipose	1.1

Table 38. Panel 2D

Tissue Name	Relative Expression(%) 2Dtm2349f_ag1539	Relative Expression(%) 2dtm2829f_ag1539
Normal Colon GENPAK 061003	2.2	37.9
83219 CC Well to Mod Diff (ODO3866)	0.2	2.7
83220 CC NAT (ODO3866)	0.2	2.7
83221 CC Gr.2 rectosigmoid (ODO3868)	0.4	7.2
83222 CC NAT (ODO3868)	0.4	3.0
83235 CC Mod Diff (ODO3920)	0.7	11.4
83236 CC NAT (ODO3920)	0.5	10.7
83237 CC Gr.2 ascend colon (ODO3921)	0.0	2.8
83238 CC NAT (ODO3921)	0.0	2.8
83241 CC from Partial Hepatectomy (ODO4309)	0.3	3.9
83242 Liver NAT (ODO4309)	0.0	0.3
87472 Colon mets to lung (OD04451-01)	0.4	7.5
87473 Lung NAT (OD04451-02)	0.2	4.3
Normal Prostate Clontech A+ 6546-1	1.7	0.0
84140 Prostate Cancer (OD04410)	1.3	10.8
84141 Prostate NAT (OD04410)	0.9	21.8
87073 Prostate Cancer (OD04720-01)	100.0	43.8
87074 Prostate NAT (OD04720-02)	0.9	19.8

Normal Lung GENPAK 061010	0.2	9.8
83239 Lung Met to Muscle (ODO4286)	0.0	0.0
83240 Muscle NAT (ODO4286)	0.6	5.4
84136 Lung Malignant Cancer (OD03126)	0.0	1.3
84137 Lung NAT (OD03126)	0.2	5.6
84871 Lung Cancer (OD04404)	0.0	0.8
84872 Lung NAT (OD04404)	0.6	5.0
84875 Lung Cancer (OD04565)	0.0	1.2
84876 Lung NAT (OD04565)	0.3	2.3
85950 Lung Cancer (OD04237-01)	0.4	6.0
85970 Lung NAT (OD04237-02)	0.0	4.9
83255 Ocular Mel Met to Liver (ODO4310)	0.0	1.4
83256 Liver NAT (ODO4310)	0.0	2.1
84139 Melanoma Mets to Lung (OD04321)	0.0	0.7
84138 Lung NAT (OD04321)	0.3	3.1
Normal Kidney GENPAK 061008	1.7	21.9
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.4	18.6
83787 Kidney NAT (OD04338)	0.6	10.5
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.6	10.1
83789 Kidney NAT (OD04339)	1.1	16.8
83790 Kidney Ca, Clear cell type (OD04340)	0.4	6.2
83791 Kidney NAT (OD04340)	0.9	11.5
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	0.0
83793 Kidney NAT (OD04348)	0.4	8.7
87474 Kidney Cancer (OD04622-01)	0.0	0.6
87475 Kidney NAT (OD04622-03)	0.0	0.8
85973 Kidney Cancer (OD04450-01)	0.2	5.0
85974 Kidney NAT (OD04450-03)	0.3	6.1
Kidney Cancer Clontech 8120607	0.2	3.5
Kidney NAT Clontech 8120608	0.4	1.1
Kidney Cancer Clontech 8120613	0.2	2.8
Kidney NAT Clontech 8120614	0.1	5.4
Kidney Cancer Clontech 9010320	0.0	1.9
Kidney NAT Clontech 9010321	0.6	8.6
Normal Uterus GENPAK 061018	0.3	1.4
Uterus Cancer GENPAK 064011	1.1	17.0
Normal Thyroid Clontech A+ 6570-1	0.8	6.8
Thyroid Cancer GENPAK 064010	0.3	4.0
Thyroid Cancer INVITROGEN A302152	0.4	7.9
Thyroid NAT INVITROGEN A302153	0.3	9.0
Normal Breast GENPAK 061019	1.2	16.0
84877 Breast Cancer (OD04566)	2.3	40.1
85975 Breast Cancer (OD04590-01)	1.2	17.8
85976 Breast Cancer Mets (OD04590-03)	1.2	12.3
87070 Breast Cancer Metastasis (OD04655-05)	1.7	23.2
GENPAK Breast Cancer 064006	0.8	15.8

Breast Cancer Res. Gen. 1024	7.5	**100.0**
Breast Cancer Clontech 9100266	0.8	7.1
Breast NAT Clontech 9100265	0.4	8.2
Breast Cancer INVITROGEN A209073	1.0	19.2
Breast NAT INVITROGEN A2090734	1.1	11.9
Normal Liver GENPAK 061009	0.0	3.8
Liver Cancer GENPAK 064003	0.2	1.2
Liver Cancer Research Genetics RNA 1025	0.0	3.7
Liver Cancer Research Genetics RNA 1026	0.0	1.4
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.6	3.0
Paired Liver Tissue Research Genetics RNA 6004-N	0.1	0.6
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	0.5
Paired Liver Tissue Research Genetics RNA 6005-N	0.0	0.3
Normal Bladder GENPAK 061001	0.2	7.7
Bladder Cancer Research Genetics RNA 1023	0.1	2.3
Bladder Cancer INVITROGEN A302173	0.2	3.0
87071 Bladder Cancer (OD04718-01)	0.0	1.3
87072 Bladder Normal Adjacent (OD04718-03)	0.9	19.1
Normal Ovary Res. Gen.	0.0	3.6
Ovarian Cancer GENPAK 064008	0.7	10.0
87492 Ovary Cancer (OD04768-07)	0.2	3.7
87493 Ovary NAT (OD04768-08)	0.2	1.9
Normal Stomach GENPAK 061017	1.2	15.4
Gastric Cancer Clontech 9060358	0.3	2.9
NAT Stomach Clontech 9060359	0.2	2.1
Gastric Cancer Clontech 9060395	0.4	8.2
NAT Stomach Clontech 9060394	0.3	4.2
Gastric Cancer Clontech 9060397	0.2	5.1
NAT Stomach Clontech 9060396	0.2	1.4
Gastric Cancer GENPAK 064005	0.2	6.8

Table 39. Panel 4.1D

Tissue Name	Relative Expression(%) 4.1x4tm6516f_ ag1539_a1	Tissue Name	Relative Expression(%) 4.1x4tm6516f_ ag1539_a1
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.5	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting	0.9	93781_HUVEC	0.0

day 4-6 in IL-2		(Endothelial)_IL-11	
93571_Secondary Tr1_resting day 4-6 in IL-2	0.6	93583_Lung Microvascular Endothelial Cells_none	0.7
		93584_Lung Microvascular Endothelial Cells_TNFa (4	
93568_primary Th1_anti-CD28/anti-CD3	0.2	ng/ml) and IL1b (1 ng/ml)	0.2
93569_primary Th2_anti-CD28/anti-CD3	0.7	92662_Microvascular Dermal endothelium_none	0.3
		92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml)	
93570_primary Tr1_anti-CD28/anti-CD3	0.0	and IL1b (1 ng/ml)	0.0
		93773_Bronchial epithelium_TNFa (4 ng/ml) and	
93565_primary Th1_resting dy 4-6 in IL-2	0.0	IL1b (1 ng/ml) **	3.6
93566_primary Th2_resting dy 4-6 in IL-2	0.0	93347_Small Airway Epithelium_none	0.7
		93348_Small Airway Epithelium_TNFa (4 ng/ml)	
93567_primary Tr1_resting dy 4-6 in IL-2	1.1	and IL1b (1 ng/ml)	0.9
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	1.4	92668_Coronery Artery SMC_resting	0.4
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	1.6	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.7
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	93107_astrocytes_resting	6.1
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.6	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	2.8
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	1.1	92666_KU-812 (Basophil)_resting	0.0
		92667_KU-812 (Basophil)_PMA/ionoycin	0.0
93354_CD4_none	2.9	93579_CCD1106 (Keratinocytes)_none	0.7
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	1.3	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.7
93103_LAK cells_resting	1.5	93791_Liver Cirrhosis	0.8
93788_LAK cells_IL-2	1.6	93577_NCI-H292	5.3
93787_LAK cells_IL-2+IL-12	0.4		
93789_LAK cells_IL-2+IFN gamma	2.1	93358_NCI-H292_IL-4	2.7
93790_LAK cells_IL-2+ IL-18	2.0	93360_NCI-H292_IL-9	5.6
93104_LAK cells_PMA/ionomycin and IL-18	0.2	93359_NCI-H292_IL-13	0.0
93578_NK Cells IL-2_resting	0.4	93357_NCI-H292_IFN gamma	0.8
93109_Mixed Lymphocyte Reaction_Two Way MLR	2.6	93777_HPAEC_-	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	2.2	93778_HPAEC_IL-1 beta/TNA alpha	0.0

93111_Mixed Lymphocyte Reaction_Two Way MLR	0.4	93254_Normal Human Lung Fibroblast_none	8.5
93112_Mononuclear Cells (PBMCs)_resting	0.5	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.3
93113_Mononuclear Cells (PBMCs)_PWM	0.0	93257_Normal Human Lung Fibroblast_IL-4	0.8
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	93256_Normal Human Lung Fibroblast_IL-9	3.1
93249_Ramos (B cell)_none	0.0	93255_Normal Human Lung Fibroblast_IL-13	0.5
93250_Ramos (B cell)_ionomycin	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	1.6
93349_B lymphocytes_PWM	0.0	93106_Dermal Fibroblasts CCD1070_resting	0.0
93350_B lymphoytes_CD40L and IL-4	1.1	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	1.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.8	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	1.3
93248_EOL-1 (Eosinophil)_dbcAMP/PMAion omycin	0.0	93772_dermal fibroblast_IFN gamma	3.3
93356_Dendritic Cells_none	0.4	93771_dermal fibroblast_IL-4	2.7
93355_Dendritic Cells_LPS 100 ng/ml	0.0	93892_Dermal fibroblasts_none	4.1
93775_Dendritic Cells_anti-CD40	0.0	99202_Neutrophils_TNFa+LPS	0.4
93774_Monocytes_resting	1.3	99203_Neutrophils_none	1.2
93776_Monocytes_LPS 50 ng/ml	0.3	735010_Colon_normal	4.4
93581_Macrophages_resting	0.3	735019_Lung_none	5.6
93582_Macrophages_LPS 100 ng/ml	0.0	64028-1_Thymus_none	25.8
93098_HUVEC (Endothelial)_none	0.0	64030-1_Kidney_none	**100.0**
93099_HUVEC (Endothelial)_starved	0.0		

Table 40. Panel CNSD.01

Tissue Name	Relative Expression(%) cns_1x4tm654 8f_ag1539_a2	Tissue Name	Relative Expression(%) cns_1x4tm654 8f_ag1539_a2
102633_BA4 Control	29.0	102605_BA17 PSP	35.0
102641_BA4 Control2	39.6	102612_BA17 PSP2	17.3
102625_BA4 Alzheimer's2	19.1	102637_Sub Nigra Control	29.8
102649_BA4 Parkinson's	69.4	102645_Sub Nigra Control2	10.3
102656_BA4 Parkinson's2	62.4	102629_Sub Nigra Alzheimer's2	10.7

102664_BA4 Huntington's	21.0	102660_Sub Nigra Parkinson's2	26.1
102671_BA4 Huntington's2	8.5	102667_Sub Nigra Huntington's	65.0
102603_BA4 PSP	19.8	102674_Sub Nigra Huntington's2	11.5
102610_BA4 PSP2	18.2	102614_Sub Nigra PSP2	0.0
102588_BA4 Depression	27.0	102592_Sub Nigra Depression	7.0
102596_BA4 Depression2	17.5	102599_Sub Nigra Depression2	5.0
102634_BA7 Control	53.0	102636_Glob Palladus Control	19.8
102642_BA7 Control2	58.2	102644_Glob Palladus Control2	12.3
102626_BA7 Alzheimer's2	18.5	102620_Glob Palladus Alzheimer's	8.8
102650_BA7 Parkinson's	35.1	102628_Glob Palladus Alzheimer's2	49.1
102657_BA7 Parkinson's2	53.0	102652_Glob Palladus Parkinson's	89.9
102665_BA7 Huntington's	72.5	102659_Glob Palladus Parkinson's2	9.6
102672_BA7 Huntington's2	34.3	102606_Glob Palladus PSP	8.2
102604_BA7 PSP	70.3	102613_Glob Palladus PSP2	4.1
102611_BA7 PSP2	30.1	102591_Glob Palladus Depression	17.4
102589_BA7 Depression	14.3	102638_Temp Pole Control	7.1
102632_BA9 Control	34.9	102646_Temp Pole Control2	75.9
102640_BA9 Control2	73.9	102622_Temp Pole Alzheimer's	9.4
102617_BA9 Alzheimer's	15.5	102630_Temp Pole Alzheimer's2	17.1
102624_BA9 Alzheimer's2	19.8	102653_Temp Pole Parkinson's	38.3
102648_BA9 Parkinson's	58.0	102661_Temp Pole Parkinson's2	38.8
102655_BA9 Parkinson's2	66.2	102668_Temp Pole Huntington's	45.6
102663_BA9 Huntington's	52.5	102607_Temp Pole PSP	14.7
102670_BA9 Huntington's2	34.9	102615_Temp Pole PSP2	21.3
102602_BA9 PSP	21.1	102600_Temp Pole Depression2	9.0
102609_BA9 PSP2	6.9	102639_Cing Gyr Control	39.0
102587_BA9 Depression	20.9	102647_Cing Gyr Control2	48.6
102595_BA9 Depression2	9.6	102623_Cing Gyr Alzheimer's	12.4
102635_BA17 Control	74.2	102631_Cing Gyr Alzheimer's2	11.1
102643_BA17 Control2	**100.0**	102654_Cing Gyr Parkinson's	18.0
102627_BA17 Alzheimer's2	23.3	102662_Cing Gyr Parkinson's2	32.8
102651_BA17 Parkinson's	82.8	102669_Cing Gyr Huntington's	81.6
102658_BA17 Parkinson's2	91.3	102676_Cing Gyr Huntington's2	23.9
102666_BA17 Huntington's	59.8	102608_Cing Gyr PSP	19.6
102673_BA17 Huntington's2	36.6	102616_Cing Gyr PSP2	7.1
102590_BA17 Depression	31.9	102594_Cing Gyr Depression	19.1
102597_BA17 Depression2	46.3	102601_Cing Gyr Depression2	14.9

Panel 1.2 Summary: The 21629632_EXT gene shows rather ubiquitous expression across the samples on this panel, with highest expression in cerebral cortex (Ctmin=25) and hippocampus. See Panel 1.3D summary for explanation.

5   **Panel 1.3D Summary:** The expression of the 21629632_EXT gene is most highly represented in the samples of brain tissue and the sample of fetal muscle. The latter profile is of particular interest in that it differs significantly from that of the adult skeletal muscle. This difference implies that this protein may function to enhance muscular growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Thus, therapeutic

10   modulation of this gene could be useful in treatment of muscular related disease. For instance treatment of weak or dystrophic muscle with the protein encoded by this gene could restore muscle mass or function. The 21629632_EXT transcript also shows highly preferential expression in brain, especially in the hippocampus and cerebral cortex where the expression is fairly high (CT = 29.5). The protein encoded by the 21629632_EXT gene appears to be a

15   GPCR, making it an excellent small molecule target. Both the hippocampus and cerebral cortex are affected by neurodegeneration in Alzheimer's disease; thus this molecule is an excellent candidate for a drug target for the treatment/prevention of Alzheimer's disease, and may also be useful for memory enhancement/processing in healthy subjects.

Panel 2D Summary: The expression profile of the 21629632_EXT gene on this panel

20   was assessed in duplicate runs, in which one run, designated as 2Dtm2349f was deemed to be erroneous. It appears that one sample of prostate cancer is contaminated with genomic DNA causing a skew in the data presentation. If this run is disregarded this gene appears to be expressed to a significant degree in a number of tissues. Particularly predominant is its expression in breast cancer and to a lesser degree in prostate cancer. Thus, therapeutic

25   modulation of this gene may be of use in the treatment of breast cancer and/or prostate cancer or other breast and/or prostate related disease.

Panel 4.1D Summary: The 21629632_EXT gene is expressed at high levels in the kidney and at somewhat lower levels in the thymus. The 21629632_EXT transcript, the protein encoded for by the transcript, or antibodies designed with the protein could be used to

30   identify kidney and thymus tissue.

Panel CNSD.01 Summary: An examination of 21629632_EXT gene expression in 8 brain regions across 12 individuals confirms that this protein is expressed in the brain of most, if not all individuals including those suffering from neurologic/psychiatric disease. Utility as a drug target would benefit from likely expression in most disease states.

I. GPCR10 (also known as 18234044_EXT)

Expression of gene 18234044_EXT was assessed using the primer-probe set Ag1539, described in Table 41. Results of the RTQ-PCR run are shown in Table 42.

Table 41. Probe name Ag1283

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TGATGGGACTCTTCAGACAATC-3 '	59.1	22	69	104
Probe	FAM-5'-AACATCCAATGGCCAATATCACCTGG-3 '-TAMRA	69.3	26	93	105
Reverse	5'-AAGAGTCCCAACAGGATGAAAT-3 '	59	22	144	106

Table 42. Panel 4.1D

Tissue Name	Relative Expression(%) 4.1dx4tm6521f _ag1283_a1	Tissue Name	Relative Expression(%) 4.1dx4tm6521f _ag1283_a1
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	93779_HUVEC (Endothelial)_IFN gamma	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	93781_HUVEC (Endothelial)_IL-11	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	93583_Lung Microvascular Endothelial Cells_none	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and ILlb (1 ng/ml)	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and ILlb (1 ng/ml)	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and ILlb (1 ng/ml) **	0.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	93347_Small Airway Epithelium_none	0.0
93567_primary Tr1_resting dy	0.0	93348_Small Airway	0.0

4-6 in IL-2

93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0
93354_CD4_none	0.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0
93103_LAK cells_resting	0.0
93788_LAK cells_IL-2	0.0
93787_LAK cells_IL-2+IL-12	0.0
93789_LAK cells_IL-2+IFN gamma	0.0
93790_LAK cells_IL-2+ IL-18	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0
93578_NK Cells IL-2_resting	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0
93249_Ramos (B cell)_none	0.0
93250_Ramos (B cell)_ionomycin	0.0
93349_B lymphocytes_PWM	0.0
93350_B lymphoytes_CD40L and IL-4	0.0

Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	
92668_Coronery Artery SMC_resting	0.0
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93107_astrocytes_resting	0.0
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
92666_KU-812 (Basophil)_resting	0.0
92667_KU-812 (Basophil)_PMA/ionoycin	0.0
93579_CCD1106 (Keratinocytes)_none	0.0
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0
93791_Liver Cirrhosis	0.0
93577_NCI-H292	0.0
93358_NCI-H292_IL-4	0.0
93360_NCI-H292_IL-9	0.0
93359_NCI-H292_IL-13	0.0
93357_NCI-H292_IFN gamma	0.0
93777_HPAEC_-	0.0
93778_HPAEC_IL-1 beta/TNA alpha	0.0
93254_Normal Human Lung Fibroblast_none	0.0
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93257_Normal Human Lung Fibroblast_IL-4	0.0
93256_Normal Human Lung Fibroblast_IL-9	0.0
93255_Normal Human Lung Fibroblast_IL-13	0.0
93258_Normal Human Lung Fibroblast_IFN gamma	0.0
93106_Dermal Fibroblasts CCD1070_resting	0.0
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0

188

92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	93772_dermal fibroblast_IFN gamma	0.9
93356_Dendritic Cells_none	0.0	93771_dermal fibroblast_IL-4	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	93892_Dermal fibroblasts_none	1.0
93775_Dendritic Cells_anti-CD40	0.0	99202_Neutrophils_TNFa+LPS	0.0
93774_Monocytes_resting	0.0	99203_Neutrophils_none	0.0
93776_Monocytes_LPS 50 ng/ml	0.0	735010_Colon_normal	0.0
93581_Macrophages_resting	0.0	735019_Lung_none	1.1
93582_Macrophages_LPS 100 ng/ml	0.0	64028-1_Thymus_none	8.4
93098_HUVEC (Endothelial)_none	0.0	64030-1_Kidney_none	100.0
93099_HUVEC (Endothelial)_starved	0.0		

Panel 2.2 Summary: Expression of the 18234044_EXT gene was low/undetectable (CT values > 35) in all samples on this panel and thus has not been shown.

Panel 4.1D Summary: The 18234044_EXT gene is expressed at high levels in the kidney. The 18234044_EXT transcript, the protein encoded for by the transcript or antibodies designed with the protein could be used to identify kidney tissue.

Example 3. SNP analysis of GPCRX clones

SeqCallingTM Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the

189

assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Variant sequences are included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a

5   "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which

10   one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in .

15   any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

Method of novel SNP Identification: SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in

20   assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2; Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth

25   of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

30   **Method of novel SNP Confirmation:** SNPs are confirmed employing a validated method know as Pyrosequencing (Pyrosequencing, Westborough, MA). Detailed protocols for Pyrosequencing can be found in: Alderborn et al. Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. (2000). *Genome Research.* 10, Issue 8, August. 1249-1265.   In brief, Pyrosequencing is a real time primer extension process

of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each dNTP upon DNA

5   chain elongation. Following Klenow polymerase-mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes proportional to the number of added bases, up to about four bases. To allow processivity of the

10   method dNTP excess is degraded by apyrase, which is also present in the starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels.

      The DNA and protein sequences for the novel single nucleotide polymorphic variants

15   are reported. Variants are reported individually but any combination of all or a select subset of variants are also included. In addition, the positions of the variant bases and the variant amino acid residues are underlined.

GPCR2 SNP

      The nucleotide sequence of variant 13374652 (bold and underlined) has a T/G single

20   nucleotide polymorphism ("SNP) as shown in Table 43. The SNP occurrence in nucleotide position 381 T->G results in a protein sequence variant in amino acid position 127 resulting in Cys to Trp.

Table 43 Variant of GPCR2 nucleotide sequence.

```
  1 ATGACCATGGAAAATTATTCTATGGCAGCTCAGTTTGTCTTAGATGGTTTAACACAGCAAGCAGAGCTCCAGCTGCCCCT
 81 CTTCCTCCTGTTCCTGGGAATCTATGTGGTCACAGTAGTGGGCAACCTGGGCATGATTCTCCTGATTGCAGTCAGCCCTC
161 TACTTCACACCCCCATGTACTATTTCCTCAGCAGCTTGTCCTTCGTCGATTTCTGCTATTCCTCTGTCATTACTCCCAAA
241 ATGCTGGTGAACTTCCTAGGAAAGAAGAATACAATCCTTTACTCTGAGTGCATGGTCCAGCTCTTTTTCTTTGTGGTCTT
321 TGTGGTGGCTGAGGGTTACCTCCTGACTGCCATGGCATATGATCGCTATGTTGCCATCTGGAGCCCACTGCTTTATAATG
401 CGATCATGTCCTCATGGGTCTGCTCACTGCTAGTGCTGGCTGCCTTCTTCTTGGGCTTTCTCTCTGCCTTGACTCATACA
481 AGTGCCATGATGAAACTGTCCTTTTGCAAATCCCACATTATCAACCATTACTTCTGTGATGTTCTTCCCCTCCTCAATCT
561 CTCCTGCTCCAACACACACCTCAATGAGCTTCTACTTTTTATCATTGCGGGGTTTAACACCTTGGTGCCCACCCTAGCTG
641 TTGCTGTCTCCTATGCCTTCATCCTCTACAGCATCCTTCACATCCGCTCCTCAGAGGGCCGGTCCAAAGCTTTTGGAACA
721 TGCAGCTCTCATCTCATGGCTGTGGTGATCTTCTTTGGGTCCATTACCTTCATGTATTTCAAGCCCCCTTCAAGTAACTC
801 CCTGGACCAGGAGAAGGTGTCCTCTGTGTTCTACACCACGGTGATCCCCATGCTGAACCCTTTAATATACAGTCTGAGGA
881 ATAAGGATGTGAAGAAAGCATTAAGGAAGGTCTTAGTAGGAAAATGA
```

GPCR4 SNP

The nucleotide sequence of SNP variant 13374653 (bold and underlined) has a C/T as shown in Table 44. The SNP occurrence in nucleotide position 348 C->T does not result in a protein sequence variant.

5

Table 44 Variant of GPCR4 nucleotide sequence.

6A1. Nucleotide sequence of variant 13374653 (underlined).

```
  1 ATGCTGGCTAGAAACAACTCCTTAGTGACTGAATTTATTCTTGCTGGATTAACAGATCATCCAGAGTTCCAGCAACCCCT
 81 CTTTTTCCTGTTTCTAGTGGTCTACATTGTCACCATGGTAGGCAACCTTGGCTTGATCATTCTTTTCGGTCTAAATTCTC
161 ACCTCCACACACCAATGTACTATTTCCTCTTCAATCTCTCCTTCATTGATCTCTGTTACTCCTCTGTTTTCACTCCCAAA
241 ATGCTAATGAACTTTGTATCAAAAAAGAATATTATCTCCTATGTTGGGTGCATGACTCAGCTGTTTTTCTTTCTCTTTTT
321 TGTCATCTCTGAATGTTACATGTTGACTTCAATGGCATATGATCGCTATGTGGCCATCTGTAATCCATTGCTGTATAAGG
401 TCACCATGTCCCATCAGGTCTGTTCTATGCTCACTTTTGCTGCTTACATAATGGGATTGGCTGGAGCCACGGCCCACACC
481 GGGTGCATGCTTAGACTCACCTTCTGCAGTGCTAATATCATCAACCATTACTTGTGTGACATACTCCCCCTCCTCCAGCT
561 TTCCTGCACCAGCACCTATGTCAACGAGGTGGTTGTTCTCATTGTTGTGGGTATTAATATCATGGTACCCAGTTGTACCA
641 TCCTCATTTCTTATGTTTTTCATTGTCACTAGCATTCTTCATATCAAATCCACTCAAGGAAGATCAAAAGCCTTCAGTACT
721 TGTAGCTCTCATGTCATTGCTCTGTCTCTGTTTTTTGGGTCAGCGGCATTCATGTATATTAAATATTCTTCTGGATCTAT
801 GGAGCAGGGAAAAGTTTCTTCTGTTTTCTACACTAATGTGGTGCCCATGCTCAATCCTCTCATCTACAGTTTGAGGAACA
881 AGGATGTCAAAGTTGCACTGAGGAAAGCTCTGATTAAAAATTCAGAGAAGAAATATATTCTAA
```

20

EQUIVALENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is

25　contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be

30　within the scope of the following claims.

WHAT IS CLAIMED IS:

1.   An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a)   a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

(b)   a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c)   an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28; and

(d)   a variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence.

2   The polypeptide of claim 1, wherein said polypeptide comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28.

3.   The polypeptide of claim 2, wherein said allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27.

4.   The polypeptide of claim 1, wherein the amino acid sequence of said variant comprises a conservative amino acid substitution.

5.   An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

(a)   a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

193

(b)   a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c)   an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

(d)   a variant of an amino acid sequence selected from the group consisting SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence;

(e)   a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence chosen from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; and

(f)   a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

6.   The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

7.   The nucleic acid molecule of claim 5, wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

8.   The nucleic acid molecule of claim 5, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27.

9.   The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

(a)   a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27;

(b)   a nucleotide sequence differing by one or more nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, provided that no more than 20% of the nucleotides differ from said nucleotide sequence;

(c)   a nucleic acid fragment of (a); and

(d)   a nucleic acid fragment of (b).

10.   The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to a nucleotide sequence chosen from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 20, 22, 24, 26 and 27, or a complement of said nucleotide sequence.

11.   The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

(a)   a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence;

(b)   an isolated second polynucleotide that is a complement of the first polynucleotide; and

(c)   a nucleic acid fragment of (a) or (b).

12.   A vector comprising the nucleic acid molecule of claim 11.

13.   The vector of claim 12, further comprising a promoter operably-linked to said nucleic acid molecule.

14.   A cell comprising the vector of claim 12.

15.   An antibody that binds immunospecifically to the polypeptide of claim 1.

16.   The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17.   The antibody of claim 15, wherein the antibody is a humanized antibody.

18.   A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a)   providing the sample;

(b)   contacting the sample with an antibody that binds immunospecifically to the polypeptide; and

(c)   determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

19.   A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a)   providing the sample;

(b)   contacting the sample with a probe that binds to said nucleic acid molecule; and

(c)   determining the presence or amount of the probe bound to said nucleic acid molecule, thereby determining the presence or amount of the nucleic acid molecule in said sample.

20.   The method of claim 19 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

21.   The method of claim 20 wherein the cell or tissue type is cancerous.

22.   A method of identifying an agent that binds to a polypeptide of claim 1, the method comprising:

(a)   contacting said polypeptide with said agent; and

(b)   determining whether said agent binds to said polypeptide.

23.   The method of claim 22 wherein the agent is a cellular receptor or a downstream effector.

24.   A method for identifying an agent that modulates the expression or activity of the polypeptide of claim 1, the method comprising:

(a)   providing a cell expressing said polypeptide;

(b)   contacting the cell with said agent, and

(c)   determining whether the agent modulates expression or activity of said polypeptide,

whereby an alteration in expression or activity of said peptide indicates said agent modulates expression or activity of said polypeptide.

25.   A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of said claim with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

26.   A method of treating or preventing a GPCRX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the polypeptide of claim 1 in an amount sufficient to treat or prevent said GPCRX-associated disorder in said subject.

27.   The method of claim 26 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

28.   The method of claim 26 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

29.   The method of claim 26, wherein said subject is a human.

30.   A method of treating or preventing a GPCRX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the nucleic acid of claim 5 in an amount sufficient to treat or prevent said GPCRX-associated disorder in said subject.

31.   The method of claim 30 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

32.   The method of claim 30 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

33.   The method of claim 30, wherein said subject is a human.

34.   A method of treating or preventing a GPCRX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the antibody of claim 15 in an amount sufficient to treat or prevent said GPCRX-associated disorder in said subject.

35.   The method of claim 34 wherein the disorder is diabetes.

36.   The method of claim 34 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

37.   The method of claim 34, wherein the subject is a human.

38.   A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically-acceptable carrier.

39.   A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically-acceptable carrier.

40.   A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically-acceptable carrier.

41.   A kit comprising in one or more containers, the pharmaceutical composition of claim 38.

42.   A kit comprising in one or more containers, the pharmaceutical composition of claim 39.

43.   A kit comprising in one or more containers, the pharmaceutical composition of claim 40.

44.   A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

(a)   measuring the level of expression of the polypeptide in a sample from the first

mammalian subject; and

(b)   comparing the amount of said polypeptide in the sample of step (a) to the amount of

the polypeptide present in a control sample from a second mammalian subject known not to

have, or not to be predisposed to, said disease;

wherein an alteration in the expression level of the polypeptide in the first subject as compared

to the control sample indicates the presence of or predisposition to said disease.

        45. The method of claim 44 wherein the predisposition is to cancers.

        46.   A method for determining the presence of or predisposition to a disease

associated with altered levels of the nucleic acid molecule of claim 5 in a first mammalian

subject, the method comprising:

(a)   measuring the amount of the nucleic acid in a sample from the first mammalian

subject; and

(b)   comparing the amount of said nucleic acid in the sample of step (a) to the amount of

the nucleic acid present in a control sample from a second mammalian subject known not to

have or not be predisposed to, the disease;

wherein an alteration in the level of the nucleic acid in the first subject as compared to the

control sample indicates the presence of or predisposition to the disease.

        47.   The method of claim 46 wherein the predisposition is to a cancer.

        48.   A method of treating a pathological state in a mammal, the method comprising

administering to the mammal a polypeptide in an amount that is sufficient to alleviate the

pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at

least 95% identical to a polypeptide comprising an amino acid sequence of at least one of SEQ

ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, or a biologically active fragment

thereof.

        49.   A method of treating a pathological state in a mammal, the method comprising

administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the

pathological state.

50.   A method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, or fragments or variants thereof, comprises the following steps:

a)   providing a polypeptide selected from the group consisting of the sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, or a peptide fragment or a variant thereof;

b)   obtaining a candidate substance;

c)   bringing into contact said polypeptide with said candidate substance; and

d)   detecting the complexes formed between said polypeptide and said candidate substance.

51.   A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein said method comprises:

a)   providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

b)   preparing membrane extracts of said recombinant eukaryotic host cell;

c)   bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and

d)   detecting the production level of second messengers metabolites.

52.   A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein said method comprises:

a)   providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group consisting of polypeptides comprising the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

b)   infecting an olfactory epithelium with said adenovirus;

c)   bringing into contact the olfactory epithelium b) with a selected ligand molecule; and

d)   detecting the increase of the response to said ligand molecule.

52.   A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28, wherein said method comprises:

a)   providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group consisting of polypeptides comprising the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 21, 23, 25 and 28;

b)   infecting an olfactory epithelium with said adenovirus;

c)   bringing into contact the olfactory epithelium b) with a selected ligand molecule; and

d)   detecting the increase of the response to said ligand molecule.

(72) Inventors; and
(75) Inventors/Applicants (for US only): SPADERNA, Steven, K. [US/US]; 261 Deerfiled Drive, Berlin, CN 06037 (US). PADIGARU, Muralidhara [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). SPYTEK, Kimberly, A. [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). CASMAN, Stacie [US/US]; 17 Peck Street, North Haven, CT 06473 (US). RASTELLI, Luca [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). MAYANKAR, Uriel [IN/US]; 229 Branford Road, Number 330, Branford, CT 06405 (US). TCHERNEV, Velizar [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US).

(54) Title: G-PROTEIN COUPLED RECEPTORS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode G-coupled protein-receptor related polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The inventio further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7	C12N15/12	C12N15/86	C12Q1/68	C07K14/72	C07K16/28
	G01N33/68	A61K31/7088	A61K38/17	A61K39/395	A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7   C07K   C12Q   G01N   A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, SEQUENCE SEARCH, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 00 21999 A (CHUGAI RES INST MOLECULAR MED ;MAEDA MASATSUGU (JP); NAKATA YASUHI) 20 April 2000 (2000-04-20) see SEQ ID NO: 24 and 28 (pp. 114-118, 129-131) page 1 -page 55; claims 1-12 ---	2,5-7, 9-17, 19-23, 30-37, 39,40, 42,43, 46,47, 49-51
X	DATABASE EMBL [Online] EBI; EVANS G A ET AL: "HTGS submission" Database accession no. AC002556 XP002202495 abstract ---	5-14,19
		-/--

X	Further documents are listed in the continuation of box C.	X	Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

5 September 2002

Date of mailing of the international search report

24 09. 2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Oderwald, H

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] EBI, Hinxton, UK; 10 December 1999 (1999-12-10) CORBY N: "Human DNA sequence from clone RP11-112J3" XP002212375 see nucleotides 178300 to 180330 abstract ---	5-7,9-14
X	DATABASE EMBL [Online] EBI, Hinxton, UK; 9 September 1999 (1999-09-09) STROTMANN J ET AL: "Small subfamily of olfactory receptor genes: structural features, expression pattern and genomic organization" XP002212376 abstract -& STROTMANN J ET AL: GENE, vol. 236, no. 2, 20 August 1999 (1999-08-20), pages 281-291, XP004178064 the whole document ---	5-7,9-14
A	STADEL J M ET AL: "Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery" TRENDS IN PHARMACOLOGICAL SCIENCES, ELSEVIER TRENDS JOURNAL, CAMBRIDGE, GB, vol. 18, no. 11, 1 November 1997 (1997-11-01), pages 430-437, XP004099345 ISSN: 0165-6147 the whole document ---	1-52
A	GAT U ET AL: "OLFACTORY RECEPTOR PROTEINS EXPRESSION, CHARACTERIZATION AND PARTIAL PURIFICATION" EUROPEAN JOURNAL OF BIOCHEMISTRY, BERLIN, DE, vol. 225, no. 3, 1994, pages 1157-1168, XP002926155 ISSN:-0014-2956 the whole document ---	1-52
P,X	WO 01 27158 A (FUCHS TANIA ;GLUSMAN GUSTAVO (IL); LANCET DORON (IL); YEDA RES & D) 19 April 2001 (2001-04-19) see SEQ ID NO: 644,832,1576 and 1764 (pp. 263,336,839,981 and 982) page 1 -page 52; claims 1-33 ---	1-25, 30-33, 39,42, 44-47, 50-52

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 01 42288 A (INCYTE GENOMICS INC ;REDDY ROOPA (US); AU YOUNG JANICE (US); BAUGH) 14 June 2001 (2001-06-14) see SEQ ID NO: 11 and 50 (pp. 137, 138 and 164 page 1 -page 79; claims 1-28 ---	1,2,5-7, 9-52
E	WO 01 68805 A (SENOMYX INC) 20 September 2001 (2001-09-20) see SEQ ID NO: 15 and 16 (pp. 86 and 87) page 1 -page 83; claims 1-124 ---	1-25, 44-47, 50-52
E	WO 01 90359 A (INCYTE GENOMICS INC ;KALLICK DEBORAH A (US); PATTERSON CHANDRA (US) 29 November 2001 (2001-11-29) see SEQ ID NO: 4 and 27 (pp. 120, 121, 139 and 140) page 1 -page 79; claims 1-90 ---	1,2,4-7, 9-52
E	WO 01 66742 A (HARLAND LEE ;WALSH RODERICK T (GB); INCYTE GENOMICS INC (US); PATT) 13 September 2001 (2001-09-13) see SEQ ID NO: 11 and 32 page 1 -page 79; claims 1-28; figures 8,9,22,23 -----	5-7, 9-37,39, 40,42-50

INTERNATIONAL SEARCH REPORT

Box I   Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

   Although claims 26-37, 48, 49 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II   Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

   see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [X] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

   1-52 partially (inventions 1 and 9)

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[X] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-52 partially

    An isolated GPCR1 polypeptide comprising amino acid sequence SEQ ID NO: 2 or 4. An isolated nucleic acid molecule comprising nucleic acid sequence SEQ ID NO: 1 or 3. A vector, a cell, an antibody. methods for determining the presence or amount, for identifying agents that bind or modulate the expression, a method for modulating the activity, methods of treating, preventing and determining the presence or predisposition to a disease, pharmaceutical compositions, kits, methods for screening of a candidate substance or a ligand using said polypeptide, nucleic acid and antibody.

2. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 5-8 (GPCR2).

3. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 9 and 10 (GPCR3).

4. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 11 and 12 (GPCR4).

5. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 13-16 (GPCR5).

6. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 17-19 (GPCR6).

7. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 20 and 21 (GPCR7).

8. Claims: 1-52 partially

    same as invention 1 but comprising SEQ ID NO: 22 and 23 (GPCR8).

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

9. Claims: 1-52 partially

     same as invention 1 but comprising SEQ ID NO: 24-26 (GPCR9).

10. Claims: 1-52 partially

     same as invention 1 but comprising SEQ ID NO: 27 and 28
(GPCR10).

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0021999	A	20-04-2000	AU	6006399 A	01-05-2000
			EP	1120426 A1	01-08-2001
			WO	0021999 A1	20-04-2000
WO 0127158	A	19-04-2001	AU	1326201 A	13-02-2001
			AU	2903701 A	23-04-2001
			AU	6118100 A	13-02-2001
			WO	0107093 A1	01-02-2001
			WO	0107094 A1	01-02-2001
			WO	0127158 A2	19-04-2001
WO 0142288	A	14-06-2001	AU	1956701 A	18-06-2001
			AU	2575601 A	18-06-2001
			EP	1235860 A2	04-09-2002
			WO	0142285 A2	14-06-2001
			WO	0142288 A2	14-06-2001
WO 0168805	A	20-09-2001	AU	4736601 A	24-09-2001
			WO	0168805 A2	20-09-2001
WO 0190359	A	29-11-2001	AU	6181401 A	03-12-2001
			AU	6472101 A	26-11-2001
			WO	0187937 A2	22-11-2001
			WO	0190359 A2	29-11-2001
WO 0166742	A	13-09-2001	AU	4001901 A	17-09-2001
			WO	0166742 A2	13-09-2001

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY8, EXPRESSED HIGHLY IN BRAIN

```
ATGACGTCCACCTGCACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCC
CTCTCCAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGGTTATC
TTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTGTTGCAGCGCAAGCCG
CAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAACCTCCTCGTCACCGACCTGCTGCAG
ATTTCGCTCGTGGCCCCCTGGGTGGTGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAAC
AGCCACTTCTGCACGGCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAAC
ACCATTGTCTTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCTACCCG
TCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACCTGGATTGTGGCCATC
CTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCAGGCTGCCTTTGATGAGCGCAATGCT
CTCTGCTCCATGATCTGGGGGGCCAGCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTC
ATCGTCATTCCACTGATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGG
AGGCAGCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCAAGGAC
TGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAGTTCCAGGATGAGAGT
GAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGCCAAGGAGGGCAGAATGGAAGCCAAG
GACGGCAGCCTGAAGGCCAAGGAAGGAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCC
AGGGGCAGCGAGGAGGTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGT
AAGGAAGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCACAGAG
GTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTTGGTGAAGACGACATC
AATTTCAGTGAGGATGACGTCGAGGCAGTGAACATCCCGGAGAGCCTCCCACCCAGTCGT
CGTAACAGCAACAGCAACCCTCCTCTGCCCAGGTGCTACCAGTGCAAAGCTGCTAAAGTG
ATCTTCATCATCATTTTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTC
CTGGCCGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAATCATC
TGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGCTACATGCACAAGACC
ATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTTCTTCTGCAAGGAAAAGCCCCCGAAA
GAAGATAGCCACCCAGACCTGCCCGGAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCT
TCCTACGATTCTGCTACTTTTCCTTGA
```

(57) Abstract: The present invention describes a newly discovered human G-protein coupled receptor and its encoding polynucleotide. Also described are expression vectors, host cells, agonists, antagonists, antisense molecules, and antibodies associated with the polynucleotide and/or polypeptide of the present invention. In addition, methods for treating, diagnosing, preventing, and screening for disorders associated with aberrant cell growth, neurological conditions, and diseases or disorders related to the brain are illustrated.

WO 02/40670 A2

LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY8,
EXPRESSED HIGHLY IN BRAIN

5

This application claims benefit to provisional application U.S. Serial
No. 60/248,285, filed November 14, 2000; to provisional application U.S. Serial No.
60/268,581, filed February 14, 2001; to provisional application U.S. Serial No.
60/308,285, filed July 27, 2001; and to provisional application U.S. Serial No.
10   60/317,166, filed September 4, 2001.

FIELD OF THE INVENTION
The present invention relates to the fields of pharmacogenomics,
diagnostics and patient therapy. More specifically, the present invention relates to
methods of diagnosing and/ or treating diseases involving the Human G-Protein
15   Coupled Receptor, HGPRBMY8.

BACKGROUND OF THE INVENTION
It is well established that many medically significant biological
processes are mediated by proteins participating in signal transduction pathways that
involve G-proteins and/or second messengers, e.g., cAMP (Lefkowitz, Nature,
20   351:353-354 (1991)). Herein these proteins are referred to as proteins participating in
pathways with G-proteins or PPG proteins. Some examples of these proteins include
the GPC receptors, such as those for adrenergic agents and dopamine (Kobilka, B. K.,
et al., PNAS, 84:46-50 (1987); Kobilka, B. K., et al., Science, 238:650-656 (1987);
Bunzow, J. R., et al., Nature, 336:783-787 (1988)), G-proteins themselves, effector
25   proteins, e.g., phospholipase C, adenylate cyclase, and phosphodiesterase, and
actuator proteins, e.g., protein kinase A and protein kinase C (Simon, M. I., et al.,
Science, 252:802-8 (1991)).
For example, in one form of signal transduction, the effect of hormone
binding is activation of an enzyme, adenylate cyclase, inside the cell. Enzyme
30   activation by hormones is dependent on the presence of the nucleotide GTP, and GTP

also influences hormone binding. A G-protein connects the hormone receptors to adenylate cyclase. G-protein was shown to exchange GTP for bound GDP when activated by hormone receptors. The GTP-carrying form then binds to an activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself,

5   returns the G-protein to its basal, inactive form. Thus, the G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector, and as a clock that controls the duration of the signal.

The membrane protein gene superfamily of G-protein coupled receptors has been characterized as having seven putative transmembrane domains.

10   The domains are believed to represent transmembrane a-helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

G-protein coupled receptors have been characterized as including these

15   seven conserved hydrophobic stretches of about 20 to 30 amino acids, connecting at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors, which bind to neuroleptic drugs, used for treating psychotic and neurological disorders. Other examples of members of this family include calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic,

20   acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins, endothelial differentiation gene-1 receptor, rhodopsins, odorant, cytomegalovirus receptors, etc.

Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that

25   are believed to stabilize functional protein structure. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 has been implicated in signal transduction.

Phosphorylation and lipidation (palmitylation or farnesylation) of cysteine residues can influence signal transduction of some G-protein coupled

30   receptors. Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxyl terminus. For several G-protein

2

coupled receptors, such as the β-adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

For some receptors, the ligand binding sites of G-protein coupled receptors are believed to comprise a hydrophilic socket formed by several G-protein

5   coupled receptors transmembrane domains, which socket is surrounded by hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is postulated to face inward and form the polar ligand-binding site. TM3 has been implicated in several G-protein coupled receptors as having a ligand-binding site, such as including the TM3 aspartate

10   residue. Additionally, TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson et al., Endoc. Rev., 10:317-331(1989)). Different G-

15   protein β-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors have been identified as an important mechanism for the regulation of G-protein coupling of some G-protein coupled receptors. G-protein coupled receptors are found in numerous sites within a mammalian host.

20   G-protein coupled receptors (GPCRs) are one of the largest receptor superfamilies known. These receptors are biologically important and malfunction of these receptors results in diseases such as Alzheimer's, Parkinson, diabetes, dwarfism, color blindness, retinal pigmentosa and asthma. GPCRs are also involved in depression, schizophrenia, sleeplessness, hypertension, anxiety, stress, renal failure

25   and in several other cardiovascular, metabolic, neural, oncology and immune disorders (F. Horn and G. Vriend, J. Mol. Med., 76: 464-468 (1998)). They have also been shown to play a role in HIV infection (Y. Feng et al., Science, 272: 872-877 (1996)). The structure of GPCRs consists of seven transmembrane helices that are connected by loops. The N-terminus is always extracellular and C-terminus is

30   intracellular. GPCRs are involved in signal transduction. The signal is received at the extracellular N-terminus side. The signal can be an endogenous ligand, a chemical moiety or light. This signal is then transduced through the membrane to the cytosolic

side where a heterotrimeric protein G-protein is activated which in turn elicits a

response (F. Horn et al., Recept. and Chann., 5: 305-314 (1998)). Ligands, agonists

and antagonists, for these GPCRs are used for therapeutic purposes.

The present invention provides a newly discovered G-protein coupled

5   receptor protein, which may be involved in cellular growth properties in brain-related

tissues based on its abundance found in the brain for this receptor. The present

invention also relates to newly identified polynucleotides, polypeptides encoded by

such polynucleotides, the use of such polynucleotides and polypeptides, as well as the

production of such polynucleotides and polypeptides. More particularly, the

10   polypeptides of the present invention are human 7-transmembrane receptors. The

invention also relates to inhibiting the action of such polypeptides.

SUMMARY OF THE INVENTION

The present invention provides a novel human member of the G-

protein coupled receptor (GPCR) family (HGPRBMY8). Based on sequence

15   homology, the protein HGPRBMY8 is a candidate GPCR. Based on its protein

sequence information, the HGPRBMY8 contains seven transmembrane domains,

which is a characteristic structural feature of GPCRs. The GPCR of this invention is

closely related to the somatostatin and GPR24 receptor families based on sequence

similarity using the BLAST program. This orphan GPCR is expressed highly in

20   brain.

It is an object of the present invention to provide an isolated

HGPRBMY8 polynucleotide as depicted in SEQ ID NO:1.

It is also an object of the present invention to provide the HGPRBMY8

polypeptide, encoded by the polynucleotide of SEQ ID NO:1 (CDS=1 to 1524) and

25   having the amino acid sequence of SEQ ID NO:2, or a functional or biologically

active portion thereof.

It is a further object of the present invention to provide compositions

comprising the HGPRBMY8 polynucleotide sequence, or a fragment thereof, or the

encoded HGPRBMY8 polypeptide, or a fragment or portion thereof. Also provided

30   by the present invention are pharmaceutical compositions comprising at least one

HGPRBMY8 polypeptide, or a functional portion thereof, wherein the compositions further comprise a pharmaceutically acceptable carrier, excipient, or diluent.

It is an object of the present invention to provide a novel, isolated, and substantially purified polynucleotide that encodes the HGPRBMY8 GPCR

5   homologue. In a particular aspect, the polynucleotide comprises the nucleotide sequence of SEQ ID NO:1. The present invention also provides a polynucleotide sequence comprising the complement of SEQ ID NO:1, or variants thereof. In addition, the present invention features polynucleotide sequences, which hybridize under conditions of moderate stringency or high stringency to the polynucleotide

10   sequence of SEQ ID NO:1.

It is an object of the present invention to further provide a nucleic acid sequence encoding the HGPRBMY8 polypeptide and an antisense of the nucleic acid sequence, as well as oligonucleotides, fragments, or portions of the nucleic acid molecule or antisense molecule. Also provided are expression vectors and host cells

15   comprising polynucleotides that encode the HGPRBMY8 polypeptide.

It is an object of the invention to provide methods for producing a polypeptide comprising the amino acid sequence depicted in SEQ ID NO:2, or a fragment thereof, comprising the steps of a) cultivating a host cell containing an expression vector containing at least a functional fragment of the polynucleotide

20   sequence encoding the HGPRBMY8 protein according to this invention under conditions suitable for the expression of the encoded polypeptide; and b) recovering the polypeptide from the host cell.

It is also an object of the invention to provide antibodies, and binding fragments thereof, which bind specifically to the HGPRBMY8 polypeptide, or an

25   epitope thereof, for use as therapeutic and diagnostic agents.

It is a further object of the invention to provide methods for screening for agents which bind to, or modulate HGPRBMY8 polypeptide, e.g., agonists and antagonists, as well as the binding molecules and/ or modulators, e.g., agonists and antagonists, particularly those that are obtained from the screening methods described.

30   It is an object of the present invention to also provide a substantially purified antagonist or inhibitor of the polypeptide of SEQ ID NO:2. In this regard,

and by way of example, a purified antibody that binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:2 is provided.

It is an object of the invention to further provide substantially purified agonists or activators of the polypeptide of SEQ ID NO:2 are further provided.

5   It is another object of the present invention to provide HGPRBMY8 nucleic acid sequences, polypeptide, peptides and antibodies for use in the diagnosis and/or screening of disorders or diseases associated with expression of the polynucleotide and its encoded polypeptide as described herein.

It is a also an object of the present invention to provide kits for 10 screening and diagnosis of disorders associated with aberrant or uncontrolled cellular development and with the expression of the polynucleotide and its encoded polypeptide as described herein.

It is an object of the present invention to further provide methods for the treatment or prevention of cancers, immune disorders, or neurological disorders 15 involving administering to an individual in need of treatment or prevention an effective amount of a purified antagonist of the HGPRBMY8 polypeptide. Due to its elevated expression in brain, the novel GPCR protein of the present invention is particularly useful in treating or preventing neurological disorders, conditions, or diseases.

20   It is an object of the present invention to also provide a method for detecting a polynucleotide that encodes a G-protein coupled receptor, preferably the HGPRBMY8 polypeptide, or homologue, or fragment thereof, in a biological sample comprising the steps of: a) hybridizing the polynucleotide, or complement of the polynucleotide sequence encoding SEQ ID NO:2 to a nucleic acid material of a 25 biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding the HGPRBMY8 polypeptide, or fragment thereof, in the biological sample. The nucleic acid material may be further amplified by the polymerase chain reaction prior to hybridization.

30   It is an object of the instant invention to provide methods and compositions to detect and diagnose alterations in the HGPRBMY8 sequence in tissues and cells as they relate to ligand response.

6

It is an object of the present invention to further provide compositions for diagnosing brain-related disorders and for diagnosing or monitoring response to HGPRBMY8 therapy in humans. In accordance with the invention, the compositions detect an alteration of the normal or wild type HGPRBMY8 sequence or its

5   expression product in a patient sample of cells or tissue.

It is an object of the present invention to provide diagnostic probes for diseases and a patient's response to therapy. The probe sequence comprises the HGPRBMY8 locus polymorphism. The probes can be constructed of nucleic acids or amino acids.

10   It is an object of the present invention to further provide antibodies, and immunoreactive portions thereof, that recognize and bind to the HGPRBMY8 protein. Such antibodies can be either polyclonal or monoclonal. Antibodies that bind to the HGPRBMY8 protein can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

15   It is also an object of the present invention to provide diagnostic kits for the determination of the nucleotide sequence of human HGPRBMY8 alleles. The kits are based on amplification-based assays, nucleic acid probe assays, protein nucleic acid probe assays, antibody assays or any combination thereof.

It is an object of the instant invention to further provide methods for

20   detecting genetic predisposition, susceptibility and response to therapy related to the brain. In accordance with the invention, the method comprises isolating a human sample, for example, blood or tissue from adults, children, embryos or fetuses, and detecting at least one alteration in the wild type HGPRBMY8 sequence, or its expression product, from the sample, wherein the alterations are indicative of genetic

25   predisposition, susceptibility or altered response to therapy related to the brain.

It is an additional object of the present invention to provide methods for making determinations as to which drug to administer, dosages, duration of treatment and the like.

Further objects, features, and advantages of the present invention will

30   be better understood upon a reading of the detailed description of the invention when considered in connection with the accompanying figures/drawings.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the full-length nucleotide sequence of cDNA clone HGPRBMY8, a human G-protein coupled receptor (SEQ ID NO:1).

Figure 2 shows the amino acid sequence (SEQ ID NO:2) from the translation of the full-length HGPRBMY8 cDNA sequence.

Figure 3 shows the 5' untranslated sequence of the orphan HGPRBMY8 (SEQ ID NO:3).

Figure 4 shows the 3' untranslated sequence of the orphan HGPRBMY8 (SEQ ID NO:4).

Figure 5 shows the predicted transmembrane region of the HGPRBMY8 protein where the predicted transmembrane regions, represented by bold-faced and underlined type, correspond to the peaks with scores above 1500.

Figures 6A- 6J show the multiple sequence alignment of the translated sequence of the orphan G-protein coupled receptor, HGPRBMY8, where the GCG (Genetics Computer Group) pileup program was used to generate the alignment with several known adrenergic and serotonin receptor sequences. The blackened areas represent identical amino acids in more than half of the listed sequences and the grey highlighted areas represent similar amino acids. As shown in Figures 6A- 6J, the sequences are aligned according to their amino acids, where: HGPRBMY8 (SEQ ID NO:2) is encoded by full length HGPRBMY8 cDNA; ACM4_CHICK (SEQ ID NO:7) represents the *Gallus gallus* (chicken)form of muscarinic acetylcholine receptor M4; YDBM_CAEEL (SEQ ID NO:8) is the *Caenorhabditis elegans* form of an orphan GPCR; 5H1A_HUMAN (SEQ ID NO:9) is the human form of the 5HT-1A receptor; 5H1A_MOUSE (SEQ ID NO:10) is the *Mus musculus* (house mouse) form of the 5HT-1A receptor; 5H1A_FUGRU (SEQ ID NO:11) represents the *Fugu rubripes* form of the 5HT-1A receptor; 5HT_LYMST (SEQ ID NO:12) is the *Lymnaea stagnalis* (great pond snail) form of the 5HT-1A receptor; A1AD_HUMAN (SEQ ID NO:13) is the human form of the alpha-1D adrenergic receptor; A1AD_MOUSE (SEQ ID NO:14) represents the mouse form of the alpha-1D adrenergic receptor (alpha 1D-adrenoceptor); Q13675 (SEQ ID NO:15) is the human form of the alpha 1C adrenergic receptor isoform 2; Q13729 (SEQ ID NO:16) represents the human form of the alpha 1C adrenergic receptor isoform 3; O60451 is

the human form of the alpha 1A adrenergic receptor isoform 4 (SEQ ID NO:17);
A1AA_RAT (SEQ ID NO:18) is the Rattus norvegicus (Norway rat) form of the
alpha-1A adrenergic receptor; O54913 (SEQ ID NO:19) is the Mus musculus (house
mouse) form of the alpha 1A-adrenergic receptor; A1AA_BOVIN (SEQ ID NO:20)
represents the Bos taurus (bovine) form of the alpha-1A adrenergic receptor;
A1AA_CANFA (SEQ ID NO:21) is the Canis familiaris (dog) form of the alpha-1A
adrenergic receptor; A1AA_RABIT (SEQ ID NO:22) represents the Oryctolagus
cuniculus (rabbit) form of the alpha-1A adrenergic receptor; A1AA_HUMAN (SEQ
ID NO:23) is the human form of the alpha-1A adrenergic receptor; A1AA_ORYLA
(SEQ ID NO:24) is the Oryzias latipes (japanese medaka) form of the alpha-1A
adrenergic receptor (MAR1); and O96716 (SEQ ID NO:25) represents the
Branchiostoma lanceolatum (amphioxus) form of the dopamine D1/beta receptor; and
O75963 (SEQ ID NO:40) is the human form of the G-protein coupled receptor RE2.

Figure 7 shows the expression profiling of the novel human orphan
GPCR, HGPRBMY8, as described in Example 3.

Figure 8 shows the brain-specific expression profiling of the novel
human orphan GPCR, HGPRBMY8, as described in Example 4.

Figure 9 shows the multiple sequence alignment of HGPRBMY8 and
other potential SNP variants (amino acid alignment). The blackened areas represent
identical amino acids and the grey highlighted areas represent similar amino acids.
As shown in Figure 9, the sequences are aligned according to their amino acids,
where: AL390879 (SEQ ID NO:41), AX148250 (SEQ ID NO:42), and AX080495
(SEQ ID NO:43) are compared to HGPRBMY8 (SEQ ID NO:2).

Figures 10A-D shows the multiple sequence alignment of
HGPRBMY8 and other potential SNP variants (nucleic acid alignment). The
blackened areas represent identical amino acids and the grey highlighted areas
represent similar amino acids. As shown in Figure 10, the sequences are aligned
according to their nucleic acids, where: AX080495 (SEQ ID NO:44); AL390879
(SEQ ID NO:45), AX148250 (SEQ ID NO:46), and are compared to HGPRBMY8
(SEQ ID NO:47).

Figure 11 shows the FACS profile of an untransfected CHO-
NFAT/CRE cell line.

Figure 12 shows that overexpression of HGPRBMY8 constitutively couples through the NFAT/CRE Response Element.

Figure 13 shows the FACS profile for the untransfected cAMP Response Element.

Figure 14 shows the overexpression of HGPRBMY8 results in coupling through the cAMP Response Element.

Figure 15A-D shows the localization of expressed HGPRBMY8 to the cell surface.

Figure 16A-D shows representative transfected CHO-NFAT/CRE cell lines with intermediate and high beta lactamase expression levels useful in screens to identify HGPRBMY8 agonists and/or antagonists.

Figure 17 shows the expression profiling of the novel human orphan GPCR, HGPRBMY8, as described in Example 8 and Table 1.

Figures 18A-B show the polynucleotide sequence (SEQ ID NO:48) and deduced amino acid sequence (SEQ ID NO:49) of the human G-protein coupled receptor, HGPRBMY8, comprising, or alternatively consisting of, one or more of the preducted polynucleotide polymorphic loci, in addition to, the encoded polypeptide polymorphic loci of the present invention for this particular protein.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

The present invention provides a novel isolated polynucleotide and encoded polypeptide, the expression of which is high in brain. This novel polypeptide is termed herein HGPRBMY8, an acronym for "Human G-Protein coupled Receptor BMY8". HGPRBMY8 is also referred to as GPCR58 and GPCR84.

Definitions

The HGPRBMY8 polypeptide (or protein) refers to the amino acid sequence of substantially purified HGPRBMY8, which may be obtained from any species, preferably mammalian, and more preferably, human, and from a variety of sources, including natural, synthetic, semi-synthetic, or recombinant. Functional fragments of the HGPRBMY8 polypeptide are also embraced by the present invention.

10

An "agonist" refers to a molecule which, when bound to the
HGPRBMY8 polypeptide, or a functional fragment thereof, increases or prolongs the
duration of the effect of the HGPRBMY8 polypeptide. Agonists may include
proteins, nucleic acids, carbohydrates, or any other molecules that bind to and
5   modulate the effect of HGPRBMY8 polypeptide. An antagonist refers to a molecule
which, when bound to the HGPRBMY8 polypeptide, or a functional fragment thereof,
decreases the amount or duration of the biological or immunological activity of
HGPRBMY8 polypeptide. "Antagonists" may include proteins, nucleic acids,
carbohydrates, antibodies, or any other molecules that decrease or reduce the effect of
10   HGPRBMY8 polypeptide.

"Nucleic acid sequence", as used herein, refers to an oligonucleotide,
nucleotide, or polynucleotide, and fragments or portions thereof, and to DNA or RNA
of genomic or synthetic origin which may be single- or double-stranded, and represent
the sense or anti-sense strand. By way of non-limiting example, fragments include
15   nucleic acid sequences that are greater than 20-60 nucleotides in length, and
preferably include fragments that are at least 70-100 nucleotides, or which are at least
1000 nucleotides or greater in length.

Similarly, "amino acid sequence" as used herein refers to an
oligopeptide, peptide, polypeptide, or protein sequence, and fragments or portions
20   thereof, and to naturally occurring or synthetic molecules. Amino acid sequence
fragments are typically from about 5 to about 30, preferably from about 5 to about 15
amino acids in length and retain the biological activity or function of the
HGPRBMY8 polypeptide.

Where "amino acid sequence" is recited herein to refer to an amino
25   acid sequence of a naturally occurring protein molecule, "amino acid sequence" and
like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid
sequence to the complete, native amino acid sequence associated with the recited
protein molecule. In addition, the terms HGPRBMY8 polypeptide and HGPRBMY8
protein are used interchangeably herein to refer to the encoded product of the
30   HGPRBMY8 nucleic acid sequence of the present invention.

A "variant" of the HGPRBMY8 polypeptide refers to an amino acid
sequence that is altered by one or more amino acids. The variant may have

11

"conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "non-conservative" changes, e.g., replacement of a glycine with a tryptophan. Minor variations may also include amino acid deletions or insertions, or

5   both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing functional biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

An "allele" or "allelic sequence" is an alternative form of the

10   HGPRBMY8 nucleic acid sequence. Alleles may result from at least one mutation in the nucleic acid sequence and may yield altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene, whether natural or recombinant, may have none, one, or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions,

15   additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding HGPRBMY8 polypeptide include nucleic acid sequences containing deletions, insertions and/or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a

20   functionally equivalent HGPRBMY8 polypeptide. Altered nucleic acid sequences may further include polymorphisms of the polynucleotide encoding the HGPRBMY8 polypeptide; such polymorphisms may or may not be readily detectable using a particular oligonucleotide probe. The encoded protein may also contain deletions, insertions, or substitutions of amino acid residues, which produce a silent change and

25   result in a functionally equivalent HGPRBMY8 protein. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of HGPRBMY8 protein is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged

30   amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and

12

valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide ("oligo") linked via an amide bond, similar to the peptide backbone of amino acid residues. PNAs typically comprise oligos of at least 5 nucleotides linked via amide bonds. PNAs may or may not terminate in positively charged amino acid residues to enhance binding affinities to DNA. Such amino acids include, for example, lysine and arginine, among others. These small molecules stop transcript elongation by binding to their complementary strand of nucleic acid (P.E. Nielsen et al., 1993, Anticancer Drug Des., 8:53-63). PNA may be pegylated to extend their lifespan in the cell where they preferentially bind to complementary single stranded DNA and RNA.

"Oligonucleotides" or "oligomers" refer to a nucleic acid sequence, preferably comprising contiguous nucleotides, of at least about 6 nucleotides to about 60 nucleotides, preferably at least about 8 to 10 nucleotides in length, more preferably at least about 12 nucleotides in length e.g., about 15 to 35 nucleotides, or about 15 to 25 nucleotides, or about 20 to 35 nucleotides, which can be typically used in PCR amplification assays, hybridization assays, or in microarrays. It will be understood that the term oligonucleotide is substantially equivalent to the terms primer, probe, or amplimer, as commonly defined in the art. It will also be appreciated by those skilled in the pertinent art that a longer oligonucleotide probe, or mixtures of probes, e.g., degenerate probes, can be used to detect longer, or more complex, nucleic acid sequences, for example, genomic DNA. In such cases, the probe may comprise at least 20-200 nucleotides, preferably, at least 30-100 nucleotides, more preferably, 50-100 nucleotides.

"Amplification" refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies, which are well known and practiced in the art (see, D.W. Dieffenbach and G.S. Dveksler, 1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

"Microarray" is an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon, or other type of membrane; filter; chip; glass slide; or any other type of suitable solid support.

The term "antisense" refers to nucleotide sequences, and compositions

5   containing nucleic acid sequences, which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense (i.e., complementary) nucleic acid molecules include PNA and may be produced by any method, including synthesis or transcription. Once introduced into a cell, the complementary

10   nucleotides combine with natural sequences produced by the cell to form duplexes, which block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "consensus" refers to the sequence that reflects the most

15   common choice of base or amino acid at each position among a series of related DNA, RNA or protein sequences. Areas of particularly good agreement often represent conserved functional domains.

A "deletion" refers to a change in either nucleotide or amino acid sequence and results in the absence of one or more nucleotides or amino acid

20   residues. By contrast, an insertion (also termed "addition") refers to a change in a nucleotide or amino acid sequence that results in the addition of one or more nucleotides or amino acid residues, as compared with the naturally occurring molecule. A substitution refers to the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids.

25   A "derivative" nucleic acid molecule refers to the chemical modification of a nucleic acid encoding, or complementary to, the encoded HGPRBMY8 polypeptide. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide, which retains the essential biological and/or functional characteristics of

30   the natural molecule. A derivative polypeptide is one, which is modified by glycosylation, pegylation, or any similar process that retains the biological and/or functional or immunological activity of the polypeptide from which it is derived.

The term "biologically active", i.e., functional, refers to a protein or polypeptide or fragment thereof having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HGPRBMY8, or any

5   oligopeptide thereof, to induce a specific immune response in appropriate animals or cells, for example, to generate antibodies, and to bind with specific antibodies.

The term "hybridization" refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between

10   two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases. The hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an anti-parallel configuration. A hybridization complex may be formed in solution (e.g., $C_o t$ or $R_o t$

15   analysis), or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., membranes, filters, chips, pins, or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been affixed).

The terms "stringency" or "stringent conditions" refer to the conditions

20   for hybridization as defined by nucleic acid composition, salt and temperature. These conditions are well known in the art and may be altered to identify and/or detect identical or related polynucleotide sequences in a sample. A variety of equivalent conditions comprising either low, moderate, or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition),

25   reaction milieu (in solution or immobilized on a solid substrate), nature of the target nucleic acid (DNA, RNA, base composition), concentration of salts and the presence or absence of other reaction components (e.g., formamide, dextran sulfate and/or polyethylene glycol) and reaction temperature (within a range of from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting

30   temperature). One or more factors may be varied to generate conditions, either low or high stringency that is different from but equivalent to the aforementioned conditions.

15

As will be understood by those of skill in the art, the stringency of hybridization may be altered in order to identify or detect identical or related polynucleotide sequences. As will be further appreciated by the skilled practitioner, the melting temperature, T_m, can be approximated by the formulas as known in the

5   art, depending on a number of parameters, such as the length of the hybrid or probe in number of nucleotides, or hybridization buffer ingredients and conditions (see, for example, T. Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982 and J. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor,

10   NY, 1989; Current Protocols in Molecular Biology, Eds. F.M. Ausubel et al., Vol. 1, "Preparation and Analysis of DNA", John Wiley and Sons, Inc., 1994-1995, Suppls. 26, 29, 35 and 42; pp. 2.10.7- 2.10.16; G.M. Wahl and S. L. Berger (1987; Methods Enzymol. 152:399-407); and A.R. Kimmel, 1987; Methods of Enzymol. 152:507-511). As a general guide, T_m decreases approximately 1°C –1.5°C with every 1%

15   decrease in sequence homology. Also, in general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridization reaction is initially performed under conditions of low stringency, followed by washes of varying, but higher stringency. Reference to hybridization stringency, e.g., high, moderate, or low stringency, typically relates to such washing conditions.

20   Thus, by way of non-limiting example, "high stringency" refers to conditions that permit hybridization of those nucleic acid sequences that form stable hybrids in 0.018M NaCl at about 65°C (i.e., if a hybrid is not stable in 0.018M NaCl at about 65°C, it will not be stable under high stringency conditions). High stringency conditions can be provided, for instance, by hybridization in 50% formamide, 5x

25   Denhardt's solution, 5xSSPE (saline sodium phosphate EDTA) (1x SSPE buffer comprises 0.15 M NaCl, 10 mM Na_2HPO_4, 1 mM EDTA), (or 1x SSC buffer containing 150 mM NaCl, 15 mM Na_3 citrate • 2 H_2O, pH 7.0), 0.2% SDS at about 42°C, followed by washing in 1x SSPE (or saline sodium citrate, SSC) and 0.1% SDS at a temperature of at least about 42°C, preferably about 55°C, more preferably about

30   65°C.

"Moderate stringency" refers, by non-limiting example, to conditions that permit hybridization in 50% formamide, 5x Denhardt's solution, 5xSSPE (or SSC), 0.2% SDS at 42°C (to about 50°C), followed by washing in 0.2x SSPE (or SSC) and 0.2% SDS at a temperature of at least about 42°C, preferably about 55°C, more
5   preferably about 65°C.

"Low stringency" refers, by non-limiting example, to conditions that permit hybridization in 10% formamide, 5x Denhardt's solution, 6xSSPE (or SSC), 0.2% SDS at 42°C, followed by washing in 1x SSPE (or SSC) and 0.2% SDS at a temperature of about 45°C, preferably about 50°C.

10   For additional stringency conditions, see T. Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). It is to be understood that the low, moderate and high stringency hybridization / washing conditions may be varied using a variety of ingredients, buffers and temperatures well known to and practiced by the skilled artisan.

15   The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total
20   complementarity exists between single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, as well as in the design and use of PNA molecules.

25   The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology, wherein complete homology is equivalent to identity. A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous". The inhibition of hybridization
30   of the completely complementary sequence to the target sequence may be examined using a hybridization assay (e.g., Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous

sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence or probe to the target sequence under conditions of low stringency. Nonetheless, conditions of low stringency do not permit non-specific binding; low stringency conditions require that the binding of two sequences

5   to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

10   Those having skill in the art will know how to determine percent identity between or among sequences using, for example, algorithms such as those based on the CLUSTALW computer program (J.D. Thompson et al., 1994, *Nucleic Acids Research*, 2(22):4673-4680), or FASTDB, (Brutlag et al., 1990, *Comp. App. Biosci.*, 6:237-245), as known in the art. Although the FASTDB algorithm typically

15   does not consider internal non-matching deletions or additions in sequences, i.e., gaps, in its calculation, this can be corrected manually to avoid an overestimation of the % identity. CLUSTALW, however, does take sequence gaps into account in its identity calculations.

            A "composition" comprising a given polynucleotide sequence refers

20   broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequence (SEQ ID NO:1) encoding HGPRBMY8 polypeptide (SEQ ID NO:2), or fragments thereof, may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be in association

25   with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be employed in an aqueous solution containing salts (e.g., NaCl), detergents or surfactants (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, and the like).

            The term "substantially purified" refers to nucleic acid sequences or

30   amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% to 85% free, and most preferably 90% or greater free from other components with which they are naturally associated.

The term "sample", or "biological sample", is meant to be interpreted in its broadest sense. A biological sample suspected of containing nucleic acid encoding HGPRBMY8 protein, or fragments thereof, or HGPRBMY8 protein itself, may comprise a body fluid, an extract from cells or tissue, chromosomes isolated

5   from a cell (e.g., a spread of metaphase chromosomes), organelle, or membrane isolated from a cell, a cell, nucleic acid such as genomic DNA (in solution or bound to a solid support such as for Southern analysis), RNA (in solution or bound to a solid support such as for Northern analysis), cDNA (in solution or bound to a solid support), a tissue, a tissue print and the like.

10   "Transformation" refers to a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being

15   transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and partial bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. Transformed cells also include those cells, which transiently express the inserted DNA or RNA for

20   limited periods of time.

The term "mimetic" refers to a molecule, the structure of which is developed from knowledge of the structure of HGPRBMY8 protein, or portions thereof, and as such, is able to effect some or all of the actions of HGPRBMY8 protein.

25   The term "portion" with regard to a protein (as in "a portion of a given protein") refers to fragments or segments of that protein. The fragments may range in size from four or five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO: 2" encompasses the full-length human HGPRBMY8

30   polypeptide, and fragments thereof.

The term "antibody" refers to intact molecules as well as fragments thereof, such as Fab, F(ab')$_2$, Fv, or Fc, which are capable of binding an epitopic or

antigenic determinant. Antibodies that bind to HGPRBMY8 polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest or prepared recombinantly for use as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the transition of RNA

5   or synthesized chemically, and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include, but are not limited to, bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), and thyroglobulin. The coupled peptide is then used to immunize the animal (e.g, a mouse, a rat, or a rabbit).

10   The term "humanized" antibody refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding capability, e.g., as described in U.S. Patent No. 5,585,089 to C.L. Queen et al.

           The term "antigenic determinant" refers to that portion of a molecule

15   that makes contact with a particular antibody (i.e., an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to an antigenic determinants. An antigenic determinant may compete with

20   the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

           The terms "specific binding" or "specifically binding" refer to the interaction between a protein or peptide and a binding molecule, such as an agonist, an antagonist, or an antibody. The interaction is dependent upon the presence of a

25   particular structure (i.e., an antigenic determinant or epitope) of the protein that is recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

30   The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:1 by Northern analysis is indicative of the presence of mRNA encoding HGPRBMY8

polypeptide (SEQ ID NO:2) in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

An alteration in the polynucleotide of SEQ ID NO:1 comprises any alteration in the sequence of the polynucleotides encoding HGPRBMY8 polypeptide,

5   including deletions, insertions, and point mutations that may be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes HGPRBMY8 polypeptide (e.g., by alterations in the pattern of restriction fragment length polymorphisms capable of hybridizing to SEQ ID NO:1), the inability of a selected fragment of SEQ ID NO:1 to

10   hybridize to a sample of genomic DNA (e.g., using allele-specific oligonucleotide probes), and improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the polynucleotide sequence encoding HGPRBMY8 polypeptide (e.g., using fluorescent *in situ* hybridization (FISH) to metaphase chromosome spreads).

15   DESCRIPTION OF THE PRESENT INVENTION

The present invention provides a novel human member of the G-protein coupled receptor (GPCR) family (HGPRBMY8). Based on sequence homology, the protein HGPRBMY8 is a novel human GPCR. This protein sequence has been predicted to contain seven transmembrane domains which is a characteristic

20   structural feature of GPCRs. HGPRBMY8 belongs to the "class A" of GPCR superfamily and is closely related to adrenergic and serotonin receptors based on sequence similarity. Class A is the largest sub-family of the GPCR superfamily. This particular orphan GPCR is expressed highly in brain.

HGPRBMY8 polypeptides and polynucleotides are useful for

25   diagnosing diseases related to over- or under- expression of HGPRBMY8 proteins by identifying mutations in the HGPRBMY8 gene using HGPRBMY8 probes, or by determining HGPRBMY8 protein or mRNA expression levels. HGPRBMY8 polypeptides are also useful for screening compounds, which affect activity or function of the protein. The invention encompasses the polynucleotide encoding the

30   HGPRBMY8 polypeptide and the use of the HGPRBMY8 polynucleotide or polypeptide, or composition thereof, in the screening, diagnosis, treatment, or

21

prevention of disorders associated with aberrant or uncontrolled cellular growth and/or function, such as neoplastic diseases (e.g., cancers and tumors), with particular regard to diseases or disorders related to the brain, e.g. neurological disorders.

Nucleic acids encoding human HGPRBMY8 according to the present invention were first identified from the human genomic data available from GenBank (Accession No: AC016468).

In one of its embodiments, the present invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:2 as shown in Figure 1. The HGPRBMY8 polypeptide is 508 amino acids in length and shares amino acid sequence homology with the GPCR RE2. The HGPRBMY8 polypeptide (SEQ ID NO:2) shares 24.3 % identity and 33.6 % similarity with over 400 amino acids of the GPCR RE2 sequence, wherein "similar" amino acids are those which have the same/ similar physical properties and in many cases, the function is conserved with similar residues. For example, amino acids Lysine and Arginine are similar; while residues such as Proline and Cysteine, which do not share any physical properties, are considered dissimilar. The HGPRBMY8 polypeptide shares 28.01% identity and 38.33% similarity with the *Fugu rubripes* 5-Hydroxytryptamine 1a-Alpha Receptor (5H1A_FUGRU; Acc. No.:O42385); 25.3% identity and 37.23% similarity with the human 5-Hydroxytryptamine 1a-Alpha Receptor (5H1A_HUMAN; Acc. No.:P08908); 27.56% idenity and 37.56% similarity with the Mus musculus 5-Hydroxytryptamine 1a-Alpha Receptor (5H1A_MOUSE; Acc. No.:Q64264, Q60956); 25.46% identity and 37.05% similarity with the *Lymnaea stagnalis* 5-hydroxytryptamine receptor (5HT_LYMST; Acc. No.:Q25414); 23.67% identity and 33.19% similarity with the Bos taurus Alpha-1A adrenergic receptor (A1AA_BOVIN; Acc. No.: P18130); 26.21% identity and 36.9% similarity with the *Canis familiaris* Alpha-1A adrenergic receptor (A1AA_CANFA; Acc. No.: O77621); 29.47% identity and 41.05% similarity with the human Alpha-1A adrenergic receptor (A1AA_HUMAN; Acc. No.: P35348); 31.65% identity and 42.29% similarity with the *Oryzias latipes* Alpha-1A adrenergic receptor (A1AA_ORYLA; Acc. No.:Q91175); 30% identity and 41.32% similarity with the *Oryctolagus cuniculus* Alpha-1A adrenergic receptor (A1AA_RABIT; Acc. No.: O02824); 24.82% identity and 34.43% similarity with the *Rattus norvegicus* Alpha-1A adrenergic receptor

22

(A1AA_RAT; Acc. No.:P43140); 29.79% identity and 41.19% similarity with the
human Alpha-1D adrenergic receptor (A1AD_HUMAN; Acc. No.: P25100); 29.2%
identity and 40.57% similarity with the Mus musculus Alpha-1D adrenergic receptor
(A1AD_MOUSE; Acc. No.:P97714, Q61619); 23.33% identity and 31.97% similarity
5   with the *Gallus gallus* muscarinic acetylcholine receptor M4 (ACM4_CHICK; Acc.
No.:P17200); 30.53% identity and 41.58% similarity with the *Mus musculus* Alpha-
1A adrenergic receptor (O54913; Acc. No.:O54913); 29.47% identity and 41.05%
similarity with the human Alpha-1A adrenergic receptor isoform 4 (O60451; Acc.
No.:O60451); 23.59% identity and 32.82% similarity with the human G-protein
10   coupled receptor RE2 (O75963; Acc. No.:O75963); 23.99% identity and 31.81%
similarity with the Branchiostoma lanceolatum dopamine D1/Beta receptor (O96716;
Acc. No.:O96716); 29.21% identity and 40.79% similarity with the human Alpha 1C
adrenergic receptor isoform 2 (Q13675; Acc. No.:Q13675); 24.87% identity and
34.52% similarity with the human Alpha 1C adrenergic receptor isoform 3 (Q13729;
15   Acc. No.:Q13729); and 21.49% identity and 32.023% similarity with the
Caenorhabditis elegans probable G protein coupled receptor F01E11.5
(YDBM_CAEEL; Acc. No.:Q19084).

Variants of the HGPRBMY8 polypeptide are also encompassed by the
present invention. A preferred HGPRBMY8 variant has at least 75 to 80%, more
20   preferably at least 85 to 90%, and even more preferably at least 90% amino acid
sequence identity to the amino acid sequence claimed herein, and which retains at
least one biological, immunological, or other functional characteristic or activity of
the HGPRBMY8 polypeptide. Most preferred is a variant having at least 95% amino
acid sequence identity to that of SEQ ID NO:2. For example, Figures 9 and 10 show
25   multiple sequence alignments of HGPRBMY8 and single nucleotide polymorphism
(SNP) variants. Highlighted are the differences in sequence.

In a preferred embodiment, polynucleotide and polypeptide
polymorphisms are shown in Figure 18A-B. The standard one-letter abbreviation for
amino acids is used to illustrate the deduced amino acid sequence. The
30   polynucleotide sequence contains a sequence of 1527 nucleotides (SEQ ID NO:48),
encoding a polypeptide of 508 amino acids (SEQ ID NO:49). The polynucleotide
polymorphic sites are represented by an "N", in bold. The polypeptide polymorphic

sites are represented by an "X", and underlined. The present invention encompasses the polynucleotide at nucleotide position 370 as being either a "T" or a "G", the polynucleotide at nucleotide position 1055 as being either a "A" or a "G", the polynucleotide at nucleotide position 1192 as being either a "G" or a "A", the

5   polynucleotide at nucleotide position 1193 as being either a "C" or a "A", and the polynucleotide at nucleotide position 1194 as being either a "T" or a "G" of Figures 18A-B (SEQ ID NO:48), in addition to any combination thereof. The present invention also encompasses the polypeptide at amino acid position 124 as being either a "Leu" or a "Val", the polypeptide at amino acid position 352 as being either a "Asp"

10   or a "Gly", and the polypeptide at amino acid position 398 as being either a "Ala" or an "Lys" of Figures 18A-B (SEQ ID NO:49).

       These polymorphisms are useful as genetic markers for any study that attempts to look for linkage between HGPRBMY8 and a disease or disease state related to this polypeptide.

15   In preferred embodiments, the following single nucleotide polymorphism polynucleotides are encompassed by the present invention:
       CACCATTGTCTTGGTGTCAGT (SEQ ID NO:50),
       CACCATTGTCGTGGTGTCAGT (SEQ ID NO:51),
       GGTGAAGATGACATGGAGTTT (SEQ ID NO:52),

20   GGTGAAGATGGCATGGAGTTT (SEQ ID NO:53),
       GTGCAAAGCTGCTAAAGTGAT (SEQ ID NO:54),
       GTGCAAAGCTACTAAAGTGAT (SEQ ID NO:55),
       TGCAAAGCTGCTAAAGTGATC (SEQ ID NO:56),
       TGCAAAGCTGATAAAGTGATC (SEQ ID NO:57)

25   GCAAAGCTGCTAAAGTGATCT (SEQ ID NO:58), and/or
       GCAAAGCTGCGAAAGTGATCT (SEQ ID NO:59). Polypeptides encoded by these polynucleotides are also provided.

       The predicted 'T' to 'G' polynucleotide polymorphism located at nucleic acid 370 of SEQ ID NO:1 is a missense mutation resulting in a change in an

30   encoding amino acid from 'L' to 'V' at amino acid position 124 of SEQ ID NO:2.

The predicted 'A' to 'G' polynucleotide polymorphism located at nucleic acid 1055 of SEQ ID NO:1 is a missense mutation resulting in a change in an encoding amino acid from 'D' to 'G' at amino acid position 352 of SEQ ID NO:2.

The predicted 'G' to 'A' polynucleotide polymorphism located at
5   nucleic acid 1192 of SEQ ID NO:1 is a missense mutation resulting in a change in an encoding amino acid from 'A' to 'T' at amino acid position 398 of SEQ ID NO:2.

The predicted 'C' to 'A' polynucleotide polymorphism located at nucleic acid 1193 of SEQ ID NO:1 is a missense mutation resulting in a change in an encoding amino acid from 'A' to 'D' at amino acid position 398 of SEQ ID NO:2.

10   The predicted 'T' to 'G' polynucleotide polymorphism located at nucleic acid 1194 of SEQ ID NO:1 is a silent mutation and does not result in a change in amino acid.

However, taken together the predicted 'G' to 'A' polynucleotide polymorphism located at nucleic acid 1192, the predicted 'C' to 'A' polynucleotide
15   polymorphism located at nucleic acid 1193, and the predicted 'T' to 'G' polynucleotide polymorphism located at nucleic acid 1194 of SEQ ID NO:1 represent a missense mutations resulting in a change in an encoding amino acid from 'A' to 'K' at amino acid position 398 of SEQ ID NO:2.

The present invention relates to isolated nucleic acid molecules
20   comprising, or alternatively, consisting of all or a portion of the variant allele of the human HGPRBMY8 G-protein coupled receptor gene (e.g., wherein reference or wildtype human HGPRBMY8 G-protein coupled receptor gene is exemplified by SEQ ID NO:1). Preferred portions are at least 10, preferably at least 20, preferably at least 40, preferably at least 100, contiguous polynucleotides comprising anyone of the
25   human HGPRBMY8 G-protein coupled receptor gene alleles described herein and exemplified in Figures 10A-D.

In one embodiment, the invention relates to a method for predicting the likelihood that an individual will have a disorder associated with the reference allele at nucleotide position 370, 1055, 1192, 1193, and/or 1194 of SEQ ID NO:1 (or
30   diagnosing or aiding in the diagnosis of such a disorder) comprising the steps of obtaining a DNA sample from an individual to be assessed and determining the nucleotide present at position 370, 1055, 1192, 1193, and/or 1194 of SEQ ID NO:1.

25

The presence of the variant allele at this position indicates that the individual has a greater likelihood of having a disorder associated therewith than an individual having the reference allele at that position, or a greater likelihood of having more severe symptoms.

5   Conversely, the invention relates to a method for predicting the likelihood that an individual will have a disorder associated with the variant allele at nucleotide position 370, 1055, 1192, 1193, and/or 1194 of SEQ ID NO:1 (or diagnosing or aiding in the diagnosis of such a disorder) comprising the steps of obtaining a DNA sample from an individual to be assessed and determining the

10   nucleotide present at position 370, 1055, 1192, 1193, and/or 1194 of SEQ ID NO:1. The presence of the variant allele at this position indicates that the individual has a greater likelihood of having a disorder associated therewith than an individual having the reference allele at that position, or a greater likelihood of having more severe symptoms.

15   The present invention further relates to isolated proteins or polypeptides comprising, or alternatively, consisting of all or a portion of the encoded variant amino acid sequence of the human HGPRBMY8 G-protein coupled receptor polypeptide (e.g., wherein reference or wildtype human HGPRBMY8 G-protein coupled receptor polypeptide is exemplified by SEQ ID NO:2). Preferred portions are

20   at least 10, preferably at least 20, preferably at least 40, preferably at least 100, contiguous polypeptides and comprises any one of the amino acid variant alleles of the human HGPRBMY8 G-protein coupled receptor polypeptide exemplified in Figures 18A-B, or a portion of SEQ ID NO:49. Alternatively, preferred portions are at least 10, preferably at least 20, preferably at least 40, preferably at least 100,

25   contiguous polypeptides and comprises any one of the amino acid reference alleles of the human HGPRBMY8 G-protein coupled receptor protein exemplified in Figures 18A-B, or a portion of SEQ ID NO:49. The invention further relates to isolated nucleic acid molecules encoding such polypeptides or proteins, as well as to antibodies that bind to such proteins or polypeptides.

30   In another embodiment, the present invention encompasses polynucleotides, which encode the HGPRBMY8 polypeptide. Accordingly, any nucleic acid sequence, which encodes the amino acid sequence of HGPRBMY8

26

polypeptide, can be used to produce recombinant molecules that express HGPRBMY8 protein. In a particular embodiment, the present invention encompasses the HGPRBMY8 polynucleotide comprising the nucleic acid sequence of SEQ ID NO:1 as shown in Figure 1. More particularly, the present invention provides the

5   HGPRBMY8 clone. More particularly, the present invention provides the HGPRBMY8 clone, deposited at the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209 on January 24, 2001 and under ATCC Accession No. PTA-2966 according to the terms of the Budapest Treaty.

         As will be appreciated by the skilled practitioner in the art, the

10   degeneracy of the genetic code results in the production of a number of nucleotide sequences encoding HGPRBMY8 polypeptide. Some of the sequences bear minimal homology to the nucleotide sequences of any known and naturally occurring gene. Accordingly, the present invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible

15   codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring HGPRBMY8, and all such variations are to be considered as being specifically disclosed.

         Although nucleotide sequences which encode HGPRBMY8

20   polypeptide and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HGPRBMY8 polypeptide under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding HGPRBMY8 polypeptide, or its derivatives, which possess a substantially different codon usage. Codons may be selected to increase the rate at

25   which expression of the peptide/polypeptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding HGPRBMY8 polypeptide, and its derivatives, without altering the encoded amino acid sequences include the production of RNA transcripts having more

30   desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The present invention also encompasses production of DNA sequences, or portions thereof, which encode the HGPRBMY8 polypeptide, and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known and practiced by those in the art. Moreover, synthetic chemistry and other known techniques may be used to introduce mutations into a sequence encoding HGPRBMY8 polypeptide, or any fragment thereof.

In preferred embodiments, the present invention encompasses a polynucleotide lacking the initiating start codon, in addition to, the resulting encoded polypeptide of HGPRBMY8. Specifically, the present invention encompasses the polynucleotide corresponding to nucleotides 4 thru 1524 of SEQ ID NO:1, and the polypeptide corresponding to amino acids 2 thru 508 of SEQ ID NO:2. Also encompassed are recombinant vectors comprising said encoding sequence, and host cells comprising said vector.

Also encompassed by the present invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequence of HGPRBMY8, such as that shown in SEQ ID NO:1, under various conditions of stringency. Hybridization conditions are typically based on the melting temperature (T_m) of the nucleic acid binding complex or probe (see, G.M. Wahl and S.L. Berger, 1987; Methods Enzymol., 152:399-407 and A.R. Kimmel, 1987; Methods of Enzymol., 152:507-511), and may be used at a defined stringency. For example, included in the present invention are sequences capable of hybridizing under moderately stringent conditions to the HGPRBMY8 sequence of SEQ ID NO:1 and other sequences which are degenerate to those which encode HGPRBMY8 polypeptide (e.g., as a non-limiting example: prewashing solution of 2X SSC, 0.5% SDS, 1.0mM EDTA, pH 8.0, and hybridization conditions of 50°C, 5XSSC, overnight.

The nucleic acid sequence encoding the HGPRBMY8 protein may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method, which may be employed, is restriction-site PCR, which utilizes universal primers to retrieve unknown sequence adjacent to a known

28

WO 02/40670

PCT/US01/43909

locus (G. Sarkar, 1993, PCR Methods Applic., 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

5

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region or sequence (T. Triglia et al., 1988, Nucleic Acids Res., 16:8186). The primers may be designed using OLIGO 4.06 Primer Analysis software (National Biosciences Inc.; Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

10

15

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome (YAC) DNA (M. Lagerstrom et al., 1991, PCR Methods Applic., 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before performing PCR. J.D. Parker et al. (1991; Nucleic Acids Res., 19:3055-3060) provide another method which may be used to retrieve unknown sequences. In addition, PCR, nested primers, and PROMOTERFINDER libraries can be used to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

20

25

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, since they will contain more sequences, which contain the 5' regions of genes. The use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic

30

libraries may be useful for extension of sequence into the 5' and 3' non-transcribed regulatory regions.

The embodiments of the present invention can be practiced using methods for DNA sequencing which are well known and generally available in the art. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical Corp.; Cleveland, OH), Taq polymerase (PE Biosystems; Gaithersburg, MD), thermostable T7 polymerase (Amersham Pharmacia Biotechnology; Piscataway, NJ), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Life Technologies (Rockville, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton; Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research; Watertown, MA) and the ABI Catalyst and 373 and 377 DNA sequencers (PE Biosystems; Gaithersburg, MD).

Commercially available capillary electrophoresis systems may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems; Gaithersburg, MD) and the entire process -- from loading of samples to computer analysis and electronic data display -- may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA, which might be present in limited amounts in a particular sample.

In another embodiment of the present invention, polynucleotide sequences or fragments thereof which encode HGPRBMY8 polypeptide, or peptides thereof, may be used in recombinant DNA molecules to direct the expression of HGPRBMY8 polypeptide product, or fragments or functional equivalents thereof, in appropriate host cells. Because of the inherent degeneracy of the genetic code, other DNA sequences, which encode substantially the same or a functionally equivalent

amino acid sequence, may be produced and these sequences may be used to clone and
express HGPRBMY8 protein.

As will be appreciated by those having skill in the art, it may be
advantageous to produce HGPRBMY8 polypeptide-encoding nucleotide sequences
5   possessing non-naturally occurring codons. For example, codons preferred by a
particular prokaryotic or eukaryotic host can be selected to increase the rate of protein
expression or to produce a recombinant RNA transcript having desirable properties,
such as a half-life which is longer than that of a transcript generated from the
naturally occurring sequence.

10   The nucleotide sequence of the present invention can be engineered
using methods generally known in the art in order to alter HGPRBMY8 polypeptide-
encoding sequences for a variety of reasons, including, but not limited to, alterations
which modify the cloning, processing, and/or expression of the gene product. DNA
shuffling by random fragmentation and PCR reassembly of gene fragments and
15   synthetic oligonucleotides may be used to engineer the nucleotide sequences. For
example, site-directed mutagenesis may be used to insert new restriction sites, alter
glycosylation patterns, change codon preference, produce splice variants, or introduce
mutations, and the like.

In another embodiment of the present invention, natural, modified, or
20   recombinant nucleic acid sequences encoding HGPRBMY8 polypeptide may be
ligated to a heterologous sequence to encode a fusion protein. For example, for
screening peptide libraries for inhibitors of HGPRBMY8 activity, it may be useful to
encode a chimeric HGPRBMY8 protein that can be recognized by a commercially
available antibody. A fusion protein may also be engineered to contain a cleavage
25   site located between the HGPRBMY8 protein-encoding sequence and the
heterologous protein sequence, so that HGPRBMY8 protein may be cleaved and
purified away from the heterologous moiety.

In another embodiment, sequences encoding HGPRBMY8 polypeptide
may be synthesized in whole, or in part, using chemical methods well known in the
30   art (see, for example, M.H. Caruthers et al., 1980, Nucl. Acids Res. Symp. Ser., 215-
223 and T. Horn et al., 1980, Nucl. Acids Res. Symp. Ser., 225-232). Alternatively,
the protein itself may be produced using chemical methods to synthesize the amino

31

acid sequence of HGPRBMY8 polypeptide, or a fragment or portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (J.Y. Roberge et al., 1995, Science, 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (PE Biosystems; Gaithersburg, MD).

      The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (e.g., T. Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., New York, NY), by reversed-phase high performance liquid chromatography, or other purification methods as are known in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, *supra*). In addition, the amino acid sequence of HGPRBMY8 polypeptide or any portion thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

      To express a biologically active HGPRBMY8 polypeptide or peptide, the nucleotide sequences encoding HGPRBMY8 polypeptide, or functional equivalents, may be inserted into an appropriate expression vector, i.e., a vector, which contains the necessary elements for the transcription and translation of the inserted coding sequence.

      Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding HGPRBMY8 polypeptide and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in J. Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y. and in F.M. Ausubel et al., 1989, *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY.

      A variety of expression vector/ host systems may be utilized to contain and express sequences encoding HGPRBMY8 polypeptide. Such expression vector/host systems include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression

32

vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) and tobacco mosaic virus (TMV)), or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or

5   animal cell systems. The host cell employed is not limiting to the present invention.

"Control elements" or "regulatory sequences" are those non-translated regions of the vector, e.g., enhancers, promoters, 5' and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector

10   system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene; La Jolla, CA) or PSPORT1 plasmid (Life Technologies; Rockville, MD), and the like, may be used. The

15   baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes), or from plant viruses (e.g., viral promoters or leader sequences), may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferred. If it is necessary to generate a cell

20   line that contains multiple copies of the sequence encoding HGPRBMY8, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected, depending upon the use intended for the expressed HGPRBMY8 product. For example, when large quantities of expressed protein are needed for the induction of

25   antibodies, vectors, which direct high level expression of fusion proteins that are readily purified, may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene; La Jolla, CA), in which the sequence encoding HGPRBMY8 polypeptide may be ligated into the vector in-frame with sequences for the amino-terminal Met

30   and the subsequent 7 residues of ß-galactosidase, so that a hybrid protein is produced; pIN vectors (see, G. Van Heeke and S.M. Schuster, 1989, J. Biol. Chem., 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to

express foreign polypeptides, as fusion proteins with glutathione S-transferase (GST).
In general, such fusion proteins are soluble and can be easily purified from lysed cells
by adsorption to glutathione-agarose beads followed by elution in the presence of free
glutathione. Proteins made in such systems may be designed to include heparin,
5   thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of
interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing
constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH
may be used. (For reviews, see F.M. Ausubel et al., *supra*, and Grant et al., 1987,
10   *Methods Enzymol.*, 153:516-544).

Should plant expression vectors be desired and used, the expression of
sequences encoding HGPRBMY8 polypeptide may be driven by any of a number of
promoters. For example, viral promoters such as the 35S and 19S promoters of
CaMV may be used alone or in combination with the omega leader sequence from
15   TMV (N. Takamatsu, 1987, EMBO J., 6:307-311). Alternatively, plant promoters
such as the small subunit of RUBISCO, or heat shock promoters, may be used (G.
Coruzzi et al., 1984, EMBO J., 3:1671-1680; R. Broglie et al., 1984, Science,
224:838-843; and J. Winter et al., 1991, Results Probl. Cell Differ. 17:85-105). These
constructs can be introduced into plant cells by direct DNA transformation or
20   pathogen-mediated transfection. Such techniques are described in a number of
generally available reviews (see, for example, S. Hobbs or L.E. Murry, In: McGraw
Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp.
191-196).

An insect system may also be used to express HGPRBMY8
25   polypeptide. For example, in one such system, *Autographa californica* nuclear
polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in
Spodoptera frugiperda cells or in *Trichoplusia* larvae. The sequences encoding
HGPRBMY8 polypeptide may be cloned into a non-essential region of the virus such
as the polyhedrin gene and placed under control of the polyhedrin promoter.
30   Successful insertion of HGPRBMY8 polypeptide will render the polyhedrin gene
inactive and produce recombinant virus lacking coat protein. The recombinant
viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia*

larvae in which the HGPRBMY8 polypeptide product may be expressed (E.K.

Engelhard et al., 1994, Proc. Nat. Acad. Sci., 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems

may be utilized. In cases where an adenovirus is used as an expression vector,

5   sequences encoding HGPRBMY8 polypeptide may be ligated into an adenovirus

transcription/ translation complex containing the late promoter and tripartite leader

sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be

used to obtain a viable virus which is capable of expressing HGPRBMY8 polypeptide

in infected host cells (J. Logan and T. Shenk, 1984, Proc. Natl. Acad. Sci., 81:3655-

10   3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV)

enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient

translation of sequences encoding HGPRBMY8 polypeptide. Such signals include

the ATG initiation codon and adjacent sequences. In cases where sequences encoding

15   HGPRBMY8 polypeptide, its initiation codon, and upstream sequences are inserted

into the appropriate expression vector, no additional transcriptional or translational

control signals may be needed. However, in cases where only coding sequence, or a

fragment thereof, is inserted, exogenous translational control signals, including the

ATG initiation codon, should be provided. Furthermore, the initiation codon should

20   be in the correct reading frame to ensure translation of the entire insert. Exogenous

translational elements and initiation codons may be of various origins, both natural

and synthetic. The efficiency of expression may be enhanced by the inclusion of

enhancers which are appropriate for the particular cell system that is used, such as

those described in the literature (D. Scharf et al., 1994, Results Probl. Cell Differ.,

25   20:125-162).

Moreover, a host cell strain may be chosen for its ability to modulate

the expression of the inserted sequences or to process the expressed protein in the

desired fashion. Such modifications of the polypeptide include, but are not limited to,

acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation.

30   Post-translational processing which cleaves a "prepro" form of the protein may also

be used to facilitate correct insertion, folding and/or function. Different host cells

having specific cellular machinery and characteristic mechanisms for such post-

35

translational activities (e.g., CHO, HeLa, MDCK, HEK293, and W138) are available

from the American Type Culture Collection (ATCC), American Type Culture

Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, and

may be chosen to ensure the correct modification and processing of the foreign

5   protein.

For long-term, high-yield production of recombinant proteins, stable

expression is preferred. For example, cell lines which stably express HGPRBMY8

protein may be transformed using expression vectors which may contain viral origins

of replication and/or endogenous expression elements and a selectable marker gene on

10   the same, or on a separate, vector. Following the introduction of the vector, cells may

be allowed to grow for 1-2 days in an enriched cell culture medium before they are

switched to selective medium. The purpose of the selectable marker is to confer

resistance to selection, and its presence allows the growth and recovery of cells,

which successfully express the introduced sequences. Resistant clones of stably

15   transformed cells may be proliferated using tissue culture techniques appropriate to

the cell type.

Any number of selection systems may be used to recover transformed

cell lines. These include, but are not limited to, the Herpes Simplex Virus thymidine

kinase (HSV TK), (M. Wigler et al., 1977, Cell, 11:223-32) and adenine

20   phosphoribosyltransferase (I. Lowy et al., 1980, Cell, 22:817-23) genes which can be

employed in tk⁻ or aprt⁻ cells, respectively. Also, anti-metabolite, antibiotic or

herbicide resistance can be used as the basis for selection; for example, dhfr, which

confers resistance to methotrexate (M. Wigler et al., 1980, Proc. Natl. Acad. Sci.,

77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-

25 .   418 (F. Colbere-Garapin et al., 1981, J. Mol. Biol., 150:1-14); and als or pat, which

confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively

(Murry, supra). Additional selectable genes have been described, for example, trpB,

which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells

to utilize histinol in place of histidine (S.C. Hartman and R.C. Mulligan, 1988, Proc.

30   Natl. Acad. Sci., 85:8047-51). Recently, the use of visible markers has gained

popularity with such markers as the anthocyanins, ß-glucuronidase and its substrate

GUS, and luciferase and its substrate luciferin, which are widely used not only to

identify transformants, but also to quantify the amount of transient or stable protein expression that is attributable to a specific vector system (C.A. Rhodes et al., 1995, *Methods Mol. Biol.*, 55:121-131).

Although the presence or absence of marker gene expression suggests
5   that the gene of interest is also present, the presence and expression of the desired gene of interest may need to be confirmed. For example, if the nucleic acid sequence encoding HGPRBMY8 polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences encoding HGPRBMY8 polypeptide can be identified by the absence of marker gene function. Alternatively, a marker gene can
10   be placed in tandem with a sequence encoding HGPRBMY8 polypeptide under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates co-expression of the tandem gene.

Alternatively, host cells, which contain the nucleic acid, sequence encoding HGPRBMY8 polypeptide and which express HGPRBMY8 polypeptide
15   product may be identified by a variety of procedures known to those having skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques, including membrane, solution, or chip based technologies, for the detection and/or quantification of nucleic acid or protein.

20   The presence of polynucleotide sequences encoding HGPRBMY8 polypeptide can be detected by DNA-DNA or DNA-RNA hybridization, or by amplification using probes or portions or fragments of polynucleotides encoding HGPRBMY8 polypeptide. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers, based on the sequences encoding HGPRBMY8
25   polypeptide, to detect transformants containing DNA or RNA encoding HGPRBMY8 polypeptide.

A wide variety of labels and conjugation techniques are known and employed by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for
30   detecting sequences related to polynucleotides encoding HGPRBMY8 polypeptide include oligo-labeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding HGPRBMY8 polypeptide,

or any portions or fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase, such as T7, T3, or SP(6) and labeled nucleotides. These procedures may

5   be conducted using a variety of commercially available kits (e.g., Amersham Pharmacia Biotech, Promega and U.S. Biochemical Corp.). Suitable reporter molecules or labels which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

10   Furthermore, in yet another embodiment, G-protein coupled receptor-encoding polynucleotide sequences can be used to purify a molecule or compound in a sample, wherein the molecule or compound specifically binds to the polynucleotide, comprising: a) combining the G-protein coupled receptor-encoding polynucleotide, or fragment thereof, under conditions to allow specific binding; b) detecting specific

15   binding between the G-protein coupled receptor-encoding polynucleotide and the molecule or compound; c) recovering the bound polynucleotide; and d) separating the polynucleotide from the molecule or compound, thereby obtaining a purified molecule or compound.

           Host cells transformed with nucleotide sequences encoding

20   HGPRBMY8 protein, or fragments thereof, may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/ or the vector used. As will be understood by those having skill in the art, expression vectors containing polynucleotides which encode HGPRBMY8 protein

25   may be designed to contain signal sequences which direct secretion of the HGPRBMY8 protein through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join nucleic acid sequences encoding HGPRBMY8 protein to nucleotide sequence encoding a polypeptide domain, which will facilitate purification of soluble proteins. Such purification facilitating domains include, but

30   are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals; protein A domains that allow purification on immobilized immunoglobulin; and the domain utilized in the FLAGS extension/

affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and HGPRBMY8 protein may be used to facilitate purification. One such expression vector provides

5   for expression of a fusion protein containing HGPRBMY8 and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography) as described by J. Porath et al., 1992, Prot. Exp. Purif., 3:263-281, while the enterokinase cleavage site provides a means for purifying from the fusion

10   protein. For a discussion of suitable vectors for fusion protein production, see D.J. Kroll et al., 1993; DNA Cell Biol., 12:441-453.

In addition to recombinant production, fragments of HGPRBMY8 polypeptide may be produced by direct peptide synthesis using solid-phase techniques (J. Merrifield, 1963, J. Am. Chem. Soc., 85:2149-2154). Protein synthesis may be

15   performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using ABI 431A Peptide Synthesizer (PE Biosystems; Gaithersburg, MD). Various fragments of HGPRBMY8 polypeptide can be chemically synthesized separately and then combined using chemical methods to produce the full-length molecule.

20   Human artificial chromosomes (HACs) may be used to deliver larger fragments of DNA than can be contained and expressed in a plasmid vector. HACs are linear microchromosomes which may contain DNA sequences of 10K to 10M in size, and contain all of the elements that are required for stable mitotic chromosome segregation and maintenance (see, J.J. Harrington et al., 1997, *Nature Genet.*, 15:345-

25   355). HACs of 6 to 10M are constructed and delivered via conventional delivery methods (e.g., liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Diagnostic Assays

A variety of protocols for detecting and measuring the expression of

30   HGPRBMY8 polypeptide using either polyclonal or monoclonal antibodies specific for the protein are known and practiced in the art. Examples include enzyme-linked

immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive with two non-interfering epitopes on the HGPRBMY8 polypeptide is preferred, but a competitive binding assay may also be employed.

5   These and other assays are described in the art as represented by the publication of R. Hampton et al., 1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN and D.E. Maddox et al., 1983; J. Exp. Med., 158:1211-1216).

This invention also relates to the use of HGPRBMY8 polynucleotides as diagnostic reagents. Detection of a mutated form of the HGPRBMY8 gene

10   associated with a dysfunction provides a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression, or altered expression of HGPRBMY8. Individuals carrying mutations in the HGPRBMY8 gene may be detected at the DNA level by a variety of techniques.

15   Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size

20   of the amplified product in comparison to the normal genotype. Hybridizing amplified DNA to labeled HGPRBMY8 polynucleotide sequences can identify point mutations. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility

25   of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., Science (1985) 230:1242. Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method. See Cotton et al., Proc. Natl. Acad. Sci., USA (1985) 85:43297-4401. In another embodiment, an array of

30   oligonucleotides probes comprising HGPRBMY8 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be

used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M. Chee et al., Science, 274:610-613, 1996).

5   The diagnostic assays offer a process for diagnosing or determining, for example, a susceptibility to infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2 through detection of a mutation in the HGPRBMY8 gene by the methods described. The invention also provides diagnostic assays for determining or monitoring susceptibility to the following conditions, diseases, or disorders: HIV infections; asthma; allergies;

10   obesity; anorexia; bulimia; ulcers; acute heart failure; hypotension; hypertension; angina pectoris; myocardial infarction; urinary retention; osteoporosis; benign prostatic hypertrophy; cancers; brain-related disorders; Parkinson's disease; neuropathic pain; immune; metabolic; cardiovascular; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia,

15   severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome; Sydenham chorea; major depressive disorder; and obsessive-compulsive disorder (OCD). Movement type diseases, disorders, or conditions may be targeted in particular since HGPRBMY8 is expressed in the caudate nucleus of the brain.

20   In addition, infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or HIV-2, as well as, conditions, diseases, or disorders such as, HIV infections; asthma; allergies; obesity; anorexia; bulimia; ulcers; acute heart failure; hypotension; hypertension; angina pectoris; myocardial infarction; urinary retention; osteoporosis; benign prostatic hypertrophy;

25   cancers; brain-related disorders; Parkinson's disease; neuropathic pain; immune; metabolic; cardiovascular; and psychotic and neurological disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome, can be diagnosed by methods comprising determining from a sample

30   derived from a subject having an abnormally decreased or increased level of HGPRBMY8 polypeptide or HGPRBMY8 mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in

41

the art for the quantification of polynucleotides, such as, for example, PCR, RT-PCR, RNAse protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as an HGPRBMY8, in a sample derived from a host are well known to those of skill in the art. Such assay

5   methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

In another of its aspects, the present invention relates to a diagnostic kit for a disease or susceptibility to a disease, particularly infections such as bacterial, fungal, protozoan and viral infections, particularly infections caused by HIV-1 or

10   HIV-2, as well as, conditions, diseases, or disorders such as, HIV infections; asthma; allergies; obesity; anorexia; bulimia; ulcers; acute heart failure; hypotension; hypertension; angina pectoris; myocardial infarction; urinary retention; osteoporosis; benign prostatic hypertrophy; cancers; brain-related disorders; Parkinson's disease; neuropathic pain; immune; metabolic; cardiovascular; and psychotic and neurological

15   disorders, including anxiety, schizophrenia, manic depression, delirium, dementia, severe mental retardation and dyskinesias, such as Huntington's disease or Gilles dela Tourett's syndrome, which comprises:

(a) an HGPRBMY8 polynucleotide, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof; or

20   (b) a nucleotide sequence complementary to that of (a); or

(c) an HGPRBMY8 polypeptide, preferably the polypeptide of SEQ ID NO: 2, or a fragment thereof; or

(d) an antibody to an HGPRBMY8 polypeptide, preferably to the polypeptide of SEQ ID NO: 2, or combinations thereof.

25   It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component and instructions are frequently included.

The GPCR polynucleotides which may be used in the diagnostic assays according to the present invention include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and

30   quantify HGPRBMY8-encoding nucleic acid expression in biopsied tissues in which expression (or under- or overexpression) of the HGPRBMY8 polynucleotide may be correlated with disease. The diagnostic assays may be used to distinguish between the

absence, presence, and excess expression of HGPRBMY8, and to monitor regulation of HGPRBMY8 polynucleotide levels during therapeutic treatment or intervention.

In a related aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding

5   HGPRBMY8 polypeptide, or closely related molecules, may be used to identify nucleic acid sequences which encode HGPRBMY8 polypeptide. The specificity of the probe, whether it is made from a highly specific region, e.g., about 8 to 10 contiguous nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or

10   amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding HGPRBMY8 polypeptide, alleles thereof, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides, most optimally 15-35

15   nucleotides, encoding the HGPRBMY8 polypeptide. The hybridization probes of this invention may be DNA or RNA and may be derived from the nucleotide sequence of SEQ ID NO:1, or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring HGPRBMY8 protein.

Methods for producing specific hybridization probes for DNA

20   encoding the HGPRBMY8 polypeptide include the cloning of a nucleic acid sequence that encodes the HGPRBMY8 polypeptide, or HGPRBMY8 derivatives, into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides.

25   Hybridization probes may be labeled by a variety of detector/ reporter groups, e.g., radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/ biotin coupling systems, and the like.

The polynucleotide sequence encoding the HGPRBMY8 polypeptide, or fragments thereof, may be used for the diagnosis of disorders associated with

30   expression of HGPRBMY8. Examples of such disorders or conditions are described for "Therapeutics". The polynucleotide sequence encoding the HGPRBMY8 polypeptide may be used in Southern or Northern analysis, dot blot, or other

membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or

chip assays utilizing fluids or tissues from patient biopsies to detect the status of, e.g.,

levels or overexpression of HGPRBMY8, or to detect altered HGPRBMY8

expression. Such qualitative or quantitative methods are well known in the art.

5   In a particular aspect, the nucleotide sequence encoding the

HGPRBMY8 polypeptide may be useful in assays that detect activation or induction

of various neoplasms or cancers, particularly those mentioned *supra*. The nucleotide

sequence encoding the HGPRBMY8 polypeptide may be labeled by standard

methods, and added to a fluid or tissue sample from a patient, under conditions

10   suitable for the formation of hybridization complexes. After a suitable incubation

period, the sample is washed and the signal is quantified and compared with a

standard value. If the amount of signal in the biopsied or extracted sample is

significantly altered from that of a comparable control sample, the nucleotide

sequence has hybridized with nucleotide sequence present in the sample, and the

15   presence of altered levels of nucleotide sequence encoding the HGPRBMY8

polypeptide in the sample indicates the presence of the associated disease. Such

assays may also be used to evaluate the efficacy of a particular therapeutic treatment

regimen in animal studies, in clinical trials, or in monitoring the treatment of an

individual patient.

20   To provide a basis for the diagnosis of disease associated with

expression of HGPRBMY8, a normal or standard profile for expression is established.

This may be accomplished by combining body fluids or cell extracts taken from

normal subjects, either animal or human, with a sequence, or a fragment thereof,

which encodes the HGPRBMY8 polypeptide, under conditions suitable for

25   hybridization or amplification. Standard hybridization may be quantified by

comparing the values obtained from normal subjects with those from an experiment

where a known amount of a substantially purified polynucleotide is used. Standard

values obtained from normal samples may be compared with values obtained from

samples from patients who are symptomatic for disease. Deviation between standard

30   and subject (patient) values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated,

hybridization assays may be repeated on a regular basis to evaluate whether the level

of expression in the patient begins to approximate that which is observed in a normal individual. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of

5   transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the

10   cancer.

Additional diagnostic uses for oligonucleotides designed from the nucleic acid sequence encoding the HGPRBMY8 polypeptide may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably comprise two

15   nucleotide sequences, one with sense orientation (5'→3') and another with antisense (3'→5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

20   Methods suitable for quantifying the expression of HGPRBMY8 include radiolabeling or biotinylating nucleotides, co-amplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (P.C. Melby et al., 1993, J. Immunol. Methods, 159:235-244; and C. Duplaa et al., 1993, Anal. Biochem., 229-236). The speed of quantifying multiple samples may be

25   accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

Therapeutic Assays

30   The HGPRBMY8 polypeptide (SEQ ID NO:2) shares homology with somatostatin-type receptors. The HGPRBMY8 protein may play a role in

45

neurological disorders, and/or in cell cycle regulation, and/or in cell signaling. The HGPRBMY8 protein may further be involved in neoplastic, cardiovascular, and immunological disorders.

In one embodiment of the present invention, the HGPRBMY8 protein

5   may play a role in neoplastic disorders. An antagonist or inhibitor of the HGPRBMY8 polypeptide may be administered to an individual to prevent or treat a neoplastic disorder. Such disorders may include, but are not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma, and particularly, cancers of the adrenal gland, bladder, bone, bone

10   marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In a related aspect, an antibody which specifically binds to HGPRBMY8 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a

15   pharmaceutical agent to cells or tissue which express the HGPRBMY8 polypeptide.

In another embodiment of the present invention, an antagonist or inhibitory agent of the HGPRBMY8 polypeptide may be administered to an individual to prevent or treat an immunological disorder. Such disorders may include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome,

20   allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis,

25   osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma.

In a preferred embodiment of the present invention, an antagonist or

30   inhibitory agent of the HGPRBMY8 polypeptide may be administered to an individual to prevent or treat a neurological disorder, particularly since HGPRBMY8 is highly expressed in the brain. Such disorders may include, but are not limited to,

46

akathesia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, Parkinson's disease, paranoid psychoses, schizophrenia, and Tourette's disorder.

5   In preferred embodiments, the HGPRBMY8 polynucleotides and polypeptides, including agonists, antagonists, and fragments thereof, are useful for modulating intracellular cAMP associated signaling pathways.

In another embodiment of the present invention, an expression vector containing the complement of the polynucleotide encoding HGPRBMY8 polypeptide

10   may be administered to an individual to treat or prevent a neoplastic disorder, including, but not limited to, the types of cancers and tumors described above.

In yet another embodiment of the present invention, an expression vector containing the complement of the polynucleotide encoding HGPRBMY8 polypeptide may be administered to an individual to treat or prevent an immune

15   disorder, including, but not limited to, the types of immune disorders described above.

In a preferred embodiment of the present invention, an expression vector containing the complement of the polynucleotide encoding HGPRBMY8 polypeptide may be administered to an individual to treat or prevent a neurological disorder, including, but not limited to, the types of disorders described above.

20   In another embodiment, the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the present invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic

25   agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of the HGPRBMY8 polypeptide of the

30   present invention may be produced using methods which are generally known in the art. For example, the HGPRBMY8 transfected CHO-NFAT/CRE cell lines of the present invention are useful for the identification of agonists and antagonists of the

47

HGPRBMY8 polypeptide. Representative uses of these cell lines would be their inclusion in a method of identifying HGPRBMY8 agonists and antagonists. Preferably, the cell lines are useful in a method for identifying a compound that modulates the biological activity of the HGPRBMY8 polypeptide, comprising the

5   steps of (a) combining a candidate modulator compound with a host cell expressing the HGPRBMY8 polypeptide having the sequence as set forth in SEQ ID NO:2; and (b) measuring an effect of the candidate modulator compound on the activity of the expressed HGPRBMY8 polypeptide. Representative vectors expressing the HGPRBMY8 polypeptide are referenced herein (e.g., pcDNA3.1 hygro ™) or

10   otherwise known in the art.

The cell lines are also useful in a method of screening for a compounds that is capable of modulating the biological activity of HGPRBMY8 polypeptide, comprising the steps of: (a) determining the biological activity of the HGPRBMY8 polypeptide in the absence of a modulator compound; (b) contacting a host cell

15   expression the HGPRBMY8 polypeptide with the modulator compound; and (c) determining the biological activity of the HGPRBMY8 polypeptide in the presence of the modulator compound; wherein a difference between the activity of the HGPRBMY8 polypeptide in the presence of the modulator compound and in the absence of the modulator compound indicates a modulating effect of the compound.

20   Additional uses for these cell lines are described herein or otherwise known in the art. In particular, purified HGPRBMY8 protein, or fragments thereof, can be used to produce antibodies, or to screen libraries of pharmaceutical agents, to identify those which specifically bind HGPRBMY8.

Antibodies specific for HGPRBMY8 polypeptide, or immunogenic

25   peptide fragments thereof, can be generated using methods that have long been known and conventionally practiced in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by an Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

30   The present invention also encompasses the polypeptide sequences that intervene between each of the predicted HGPRBMY8 transmembrane domains. Since these regions are solvent accessible either extracellularly or intracellularly, they

48

are particularly useful for designing antibodies specific to each region. Such antibodies may be useful as antagonists or agonists of the HGPRBMY8 full-length polypeptide and may modulate its activity.

The following serve as non-limiting examples of peptides or fragments that may be used to generate antibodies:

5

MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRST	(SEQ ID NO:26)
QRKPQLLQVTNRF	(SEQ ID NO:27)
WPLNS	(SEQ ID NO:28)
DRYLSIIHPLSYPSKMTQRR	(SEQ ID NO:29)
GQAAFDERNALCSMIWGASPSYT	(SEQ ID NO:30)

10

CAARRQHALLYNVKRHSLEVRVKDCVENEDEEGAEKKEEFQDESEFRRQ
HEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEAGSEEVRESSTVA
SDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDI
NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAK   (SEQ ID
NO:31)

15

AVLAVWVDVETQVPQ	(SEQ ID NO:32)
YGYMHKTIKKEIQDMLKKFFCKEKPPKEDSHPDLPGTEGGTEGKIVPSYD	
SATFP	(SEQ ID NO:33)

The present invention also encompasses the polypeptide sequences that

20

intervene between each of the predicted HGPRBMY8 transmembrane domains. Since these regions are solvent accessible either extracellularly or intracellularly, they are particularly useful for designing antibodies specific to each region. Such antibodies may be useful as antagonists or agonists of the HGPRBMY8 full-length polypeptide and may modulate its activity.

25

In preferred embodiments, the following N-terminal HGPRBMY8 TM1-2 intertransmembrane domain deletion polypeptides are encompassed by the present invention: Q1-F13, R2-F13, K3-F13, P4-F13, Q5-F13, L6-F13, and/or L7-F13 of SEQ ID NO:27. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these N-terminal

30

HGPRBMY8 TM1-2 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

49

In preferred embodiments, the following C-terminal HGPRBMY8 TM1-2 intertransmembrane domain deletion polypeptides are encompassed by the present invention: Q1-F13, Q1-R12, Q1-N11, Q1-T10, Q1-V9, Q1-Q8, and/or Q1-L7 of SEQ ID NO:27. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these C-terminal HGPRBMY8 TM1-2 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

In preferred embodiments, the following N-terminal HGPRBMY8 TM3-4 intertransmembrane domain deletion polypeptides are encompassed by the present invention: D1-R20, R2-R20, Y3-R20, L4-R20, S5-R20, I6-R20, I7-R20, H8-R20, P9-R20, L10-R20, S11-R20, Y12-R20, P13-R20, and/or S14-R20 of SEQ ID NO:29. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these N-terminal HGPRBMY8 TM3-4 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

In preferred embodiments, the following C-terminal HGPRBMY8 TM3-4 intertransmembrane domain deletion polypeptides are encompassed by the present invention: D1-R20, D1-R19, D1-Q18, D1-T17, D1-M16, D1-K15, D1-S14, D1-P13, D1-Y12, D1-S11, D1-L10, D1-P9, D1-H8, and/or D1-I7 of SEQ ID NO:29. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these C-terminal HGPRBMY8 TM3-4 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

In preferred embodiments, the following N-terminal HGPRBMY8 TM4-5 intertransmembrane domain deletion polypeptides are encompassed by the present invention: G1-T23, Q2-T23, A3-T23, A4-T23, F5-T23, D6-T23, E7-T23, R8-T23, N9-T23, A10-T23, L11-T23, C12-T23, S13-T23, M14-T23, I15-T23, W16-T23, and/or G17-T23 of SEQ ID NO:30. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these N-terminal HGPRBMY8 TM4-5 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

50

In preferred embodiments, the following C-terminal HGPRBMY8 TM4-5 intertransmembrane domain deletion polypeptides are encompassed by the present invention: G1-T23, G1-Y22, G1-S21, G1-P20, G1-S19, G1-A18, G1-G17, G1-W16, G1-I15, G1-M14, G1-S13, G1-C12, G1-L11, G1-A10, G1-N9, G1-R8,

5   and/or G1-E7 of SEQ ID NO:30. Polynucleotide sequences encoding these polypeptides are also provided. The present invention also encompasses the use of these C-terminal HGPRBMY8 TM4-5 intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

10   In preferred embodiments, the following N-terminal HGPRBMY8 TM5-6 intertransmembrane domain deletion polypeptides are encompassed by the present invention: C1-K182, A2-K182, A3-K182, R4-K182, R5-K182, Q6-K182, H7-K182, A8-K182, L9-K182, L10-K182, Y11-K182, N12-K182, V13-K182, K14-K182, R15-K182, H16-K182, S17-K182, L18-K182, E19-K182, V20-K182, R21-

15   K182, V22-K182, K23-K182, D24-K182, C25-K182, V26-K182, E27-K182, N28-K182, E29-K182, D30-K182, E31-K182, E32-K182, G33-K182, A34-K182, E35-K182, K36-K182, K37-K182, E38-K182, E39-K182, F40-K182, Q41-K182, D42-K182, E43-K182, S44-K182, E45-K182, F46-K182, R47-K182, R48-K182, Q49-K182, H50-K182, E51-K182, G52-K182, E53-K182, V54-K182, K55-K182, A56-

20   K182, K57-K182, E58-K182, G59-K182, R60-K182, M61-K182, E62-K182, A63-K182, K64-K182, D65-K182, G66-K182, S67-K182, L68-K182, K69-K182, A70-K182, K71-K182, E72-K182, G73-K182, S74-K182, T75-K182, G76-K182, T77-K182, S78-K182, E79-K182, S80-K182, S81-K182, V82-K182, E83-K182, A84-K182, G85-K182, S86-K182, E87-K182, E88-K182, V89-K182, R90-K182, E91-

25   K182, S92-K182, S93-K182, T94-K182, V95-K182, A96-K182, S97-K182, D98-K182, G99-K182, S100-K182, M101-K182, E102-K182, G103-K182, K104-K182, E105-K182, G106-K182, S107-K182, T108-K182, K109-K182, V110-K182, E111-K182, E112-K182, N113-K182, S114-K182, M115-K182, K116-K182, A117-K182, D118-K182, K119-K182, G120-K182, R121-K182, T122-K182, E123-K182, V124-

30   K182, N125-K182, Q126-K182, C127-K182, S128-K182, I129-K182, D130-K182, L131-K182, G132-K182, E133-K182, D134-K182, D135-K182, M136-K182, E137-K182, F138-K182, G139-K182, E140-K182, D141-K182, D142-K182, I143-K182,

51

N144-K182, F145-K182, S146-K182, E147-K182, D148-K182, D149-K182, V150-

K182, E151-K182, A152-K182, V153-K182, N154-K182, I155-K182, P156-K182,

E157-K182, S158-K182, L159-K182, P160-K182, P161-K182, S162-K182, R163-

K182, R164-K182, N165-K182, S166-K182, N167-K182, S168-K182, N169-K182,

5   P170-K182, P171-K182, L172-K182, P173-K182, R174-K182, C175-K182, and/or

Y176-K182 of SEQ ID NO:31. Polynucleotide sequences encoding these

polypeptides are also provided. The present invention also encompasses the use of

these N-terminal HGPRBMY8 TM5-6 intertransmembrane domain deletion

polypeptides as immunogenic and/or antigenic epitopes as described elsewhere

10   herein.

In preferred embodiments, the following C-terminal HGPRBMY8 TM5-6

intertransmembrane domain deletion polypeptides are encompassed by the present

invention: C1-K182, C1-A181, C1-A180, C1-K179, C1-C178, C1-Q177, C1-Y176,

C1-C175, C1-R174, C1-P173, C1-L172, C1-P171, C1-P170, C1-N169, C1-S168, C1-

15   N167, C1-S166, C1-N165, C1-R164, C1-R163, C1-S162, C1-P161, C1-P160, C1-

L159, C1-S158, C1-E157, C1-P156, C1-I155, C1-N154, C1-V153, C1-A152, C1-

E151, C1-V150, C1-D149, C1-D148, C1-E147, C1-S146, C1-F145, C1-N144, C1-

I143, C1-D142, C1-D141, C1-E140, C1-G139, C1-F138, C1-E137, C1-M136, C1-

D135, C1-D134, C1-E133, C1-G132, C1-L131, C1-D130, C1-I129, C1-S128, C1-

20   C127, C1-Q126, C1-N125, C1-V124, C1-E123, C1-T122, C1-R121, C1-G120, C1-

K119, C1-D118, C1-A117, C1-K116, C1-M115, C1-S114, C1-N113, C1-E112, C1-

E111, C1-V110, C1-K109, C1-T108, C1-S107, C1-G106, C1-E105, C1-K104, C1-

G103, C1-E102, C1-M101, C1-S100, C1-G99, C1-D98, C1-S97, C1-A96, C1-V95,

C1-T94, C1-S93, C1-S92, C1-E91, C1-R90, C1-V89, C1-E88, C1-E87, C1-S86, C1-

25   G85, C1-A84, C1-E83, C1-V82, C1-S81, C1-S80, C1-E79, C1-S78, C1-T77, C1-

G76, C1-T75, C1-S74, C1-G73, C1-E72, C1-K71, C1-A70, C1-K69, C1-L68, C1-

S67, C1-G66, C1-D65, C1-K64, C1-A63, C1-E62, C1-M61, C1-R60, C1-G59, C1-

E58, C1-K57, C1-A56, C1-K55, C1-V54, C1-E53, C1-G52, C1-E51, C1-H50, C1-

Q49, C1-R48, C1-R47, C1-F46, C1-E45, C1-S44, C1-E43, C1-D42, C1-Q41, C1-

30   F40, C1-E39, C1-E38, C1-K37, C1-K36, C1-E35, C1-A34, C1-G33, C1-E32, C1-

E31, C1-D30, C1-E29, C1-N28, C1-E27, C1-V26, C1-C25, C1-D24, C1-K23, C1-

V22, C1-R21, C1-V20, C1-E19, C1-L18, C1-S17, C1-H16, C1-R15, C1-K14, C1-

V13, C1-N12, C1-Y11, C1-L10, C1-L9, C1-A8, and/or C1-H7 of SEQ ID NO:31.
Polynucleotide sequences encoding these polypeptides are also provided. The present
invention also encompasses the use of these C-terminal HGPRBMY8 TM5-6
intertransmembrane domain deletion polypeptides as immunogenic and/or antigenic

5   epitopes as described elsewhere herein.

In preferred embodiments, the following N-terminal HGPRBMY8 TM6-7
intertransmembrane domain deletion polypeptides are encompassed by the present
invention: A1-Q15, V2-Q15, L3-Q15, A4-Q15, V5-Q15, W6-Q15, V7-Q15, D8-Q15,
and/or V9-Q15 of SEQ ID NO:32. Polynucleotide sequences encoding these

10   polypeptides are also provided. The present invention also encompasses the use of
these N-terminal HGPRBMY8 TM6-7 intertransmembrane domain deletion
polypeptides as immunogenic and/or antigenic epitopes as described elsewhere
herein.

In preferred embodiments, the following C-terminal HGPRBMY8 TM6-7

15   intertransmembrane domain deletion polypeptides are encompassed by the present
invention: A1-Q15, A1-P14, A1-V13, A1-Q12, A1-T11, A1-E10, A1-V9, A1-D8,
and/or A1-V7 of SEQ ID NO:32. Polynucleotide sequences encoding these
polypeptides are also provided. The present invention also encompasses the use of
these C-terminal HGPRBMY8 TM6-7 intertransmembrane domain deletion

20   polypeptides as immunogenic and/or antigenic epitopes as described elsewhere
herein.

The HGPRBMY8 polypeptides of the present invention were determined to
comprise several phosphorylation sites based upon the Motif algorithm (Genetics
Computer Group, Inc.). The phosphorylation of such sites may regulate some

25   biological activity of the HGPRBMY8 polypeptide. For example, phosphorylation at
specific sites may be involved in regulating the proteins ability to associate or bind to
other molecules (e.g., proteins, ligands, substrates, DNA, etc.). In the present case,
phosphorylation may modulate the ability of the HGPRBMY8 polypeptide to
associate with other polypeptides, particularly cognate ligand for HGPRBMY8, or its

30   ability to modulate certain cellular signal pathways.

The HGPRBMY8 polypeptide was predicted to comprise eight PKC phosphorylation sites using the Motif algorithm (Genetics Computer Group, Inc.). In vivo, protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues. The PKC phosphorylation sites have the following consensus

5   pattern: [ST]-x-[RK], where S or T represents the site of phosphorylation and 'x' an intervening amino acid residue. Additional information regarding PKC phosphorylation sites can be found in Woodget J.R., Gould K.L., Hunter T., Eur. J. Biochem. 161:177-184(1986), and Kishimoto A., Nishiyama K., Nakanishi H., Uratsuji Y., Nomura H., Takeyama Y., Nishizuka Y., J. Biol. Chem. 260:12492-

10   12499(1985); which are hereby incorporated by reference herein.

In preferred embodiments, the following PKC phosphorylation site polypeptides are encompassed by the present invention: STCTNSTRESNSS (SEQ ID NO:76), QLLQVTNRFIFNL (SEQ ID NO:77), YPSKMTQRRGYLL (SEQ ID NO:78), EAKDGSLKAKEGS (SEQ ID NO:79), EGKEGSTKVEENS (SEQ ID

15   NO:80), KVEENSMKADKGR (SEQ ID NO:81), ESLPPSRRNSNSN (SEQ ID NO:82), and/or GYMHKTIKKEIQD (SEQ ID NO:83). Polynucleotides encoding these polypeptides are also provided. The present invention also encompasses the use of the HGPRBMY8 PKC phosphorylation site polypeptides as immunogenic and/or antigenic epitopes as described elsewhere herein.

20   The HGPRBMY8 polypeptide was predicted to comprise five casein kinase II phosphorylation sites using the Motif algorithm (Genetics Computer Group, Inc.). Casein kinase II (CK-2) is a protein serine/threonine kinase whose activity is independent of cyclic nucleotides and calcium. CK-2 phosphorylates many different proteins. The substrate specificity [1] of this enzyme can be summarized as follows:

25   (1) Under comparable conditions Ser is favored over Thr.; (2) An acidic residue (either Asp or Glu) must be present three residues from the C-terminal of the phosphate acceptor site; (3) Additional acidic residues in positions +1, +2, +4, and +5 increase the phosphorylation rate. Most physiological substrates have at least one acidic residue in these positions; (4) Asp is preferred to Glu as the provider of acidic

54

determinants; and (5) A basic residue at the N-terminal of the acceptor site decreases the phosphorylation rate, while an acidic one will increase it.

A consensus pattern for casein kinase II phosphorylations site is as follows: [ST]-x(2)-[DE], wherein 'x' represents any amino acid, and S or T is the phosphorylation site.

Additional information specific to aminoacyl-transfer RNA synthetases class-II domains may be found in reference to the following publication: Pinna L.A., Biochim. Biophys. Acta 1054:267-284(1990); which is hereby incorporated herein in its entirety.

In preferred embodiments, the following casein kinase II phosphorylation site polypeptide is encompassed by the present invention: STCTNSTRESNSSH (SEQ ID NO:84), TGTSESSVEARGSE (SEQ ID NO:85), GKEGSTKVEENSMK (SEQ ID NO:86), DDINFSEDDVEAVN (SEQ ID NO:87), and/or PPKEDSHPDLPGTE (SEQ ID NO:88). Polynucleotides encoding these polypeptides are also provided. The present invention also encompasses the use of this casein kinase II phosphorylation site polypeptide as an immunogenic and/or antigenic epitope as described elsewhere herein.

The HGPRBMY8 polypeptide was predicted to comprise two cAMP- and cGMP-dependent protein kinase phosphorylation site using the Motif algorithm (Genetics Computer Group, Inc.). There has been a number of studies relative to the specificity of cAMP- and cGMP-dependent protein kinases. Both types of kinases appear to share a preference for the phosphorylation of serine or threonine residues found close to at least two consecutive N-terminal basic residues.

A consensus pattern for cAMP- and cGMP-dependent protein kinase phosphorylation sites is as follows: [RK](2)-x-[ST], wherein "x" represents any amino acid, and S or T is the phosphorylation site.

Additional information specific to cAMP- and cGMP-dependent protein kinase phosphorylation sites may be found in reference to the following publication:

Fremisco J.R., Glass D.B., Krebs E.G, J. Biol. Chem. 255:4240-4245(1980); Glass
D.B., Smith S.B., J. Biol. Chem. 258:14797-14803(1983); and Glass D.B., El-
Maghrabi M.R., Pilkis S.J., J. Biol. Chem. 261:2987-2993(1986); which is hereby
incorporated herein in its entirety.

5   In preferred embodiments, the following cAMP- and cGMP-dependent protein
kinase phosphorylation site polypeptide is encompassed by the present invention:
LLYNVKRHSLEVRV (SEQ ID NO:89), and/or SLPPSRRNSNSNPP (SEQ ID
NO:90). Polynucleotides encoding this polypeptide are also provided. The present
invention also encompasses the use of this cAMP- and cGMP-dependent protein
10   kinase phosphorylation site polypeptide as an immunogenic and/or antigenic epitope
as described elsewhere herein.

        The HGPRBMY8 polypeptide has been shown to comprise three
glycosylation sites according to the Motif algorithm (Genetics Computer Group, Inc.).
As discussed more specifically herein, protein glycosylation is thought to serve a
15   variety of functions including: augmentation of protein folding, inhibition of protein
aggregation, regulation of intracellular trafficking to organelles, increasing resistance
to proteolysis, modulation of protein antigenicity, and mediation of intercellular
adhesion.

        Asparagine glycosylation sites have the following consensus pattern, N-{P}-
20   [ST]-{P}, wherein N represents the glycosylation site. However, it is well known that
that potential N-glycosylation sites are specific to the consensus sequence Asn-Xaa-
Ser/Thr. However, the presence of the consensus tripeptide is not sufficient to
conclude that an asparagine residue is glycosylated, due to the fact that the folding of
the protein plays an important role in the regulation of N-glycosylation. It has been
25   shown that the presence of proline between Asn and Ser/Thr will inhibit N-
glycosylation; this has been confirmed by a recent statistical analysis of glycosylation
sites, which also shows that about 50% of the sites that have a proline C-terminal to
Ser/Thr are not glycosylated. Additional information relating to asparagine
glycosylation may be found in reference to the following publications, which are
30   hereby incorporated by reference herein: Marshall R.D., Annu. Rev. Biochem.

41:673-702(1972); Pless D.D., Lennarz W.J., Proc. Natl. Acad. Sci. U.S.A. 74:134-
138(1977); Bause E., Biochem. J. 209:331-336(1983); Gavel Y., von Heijne G.,
Protein Eng. 3:433-442(1990); and Miletich J.P., Broze G.J. Jr., J. Biol. Chem.
265:11397-11404(1990).

5   In preferred embodiments, the following asparagine glycosylation site
polypeptides are encompassed by the present invention: TSTCTNSTRESNSS (SEQ
ID NO:91), STRESNSSHTCMPL (SEQ ID NO:92), and/or GEDDINFSEDDVEA
(SEQ ID NO:93). Polynucleotides encoding these polypeptides are also provided.
The present invention also encompasses the use of these HGPRBMY8 asparagine
10   glycosylation site polypeptide as immunogenic and/or antigenic epitopes as described
elsewhere herein.

         The HGPRBMY8 polypeptide was predicted to comprise eight N-
myristoylation sites using the Motif algorithm (Genetics Computer Group, Inc.). An
appreciable number of eukaryotic proteins are acylated by the covalent addition of
15   myristate (a C14-saturated fatty acid) to their N-terminal residue via an amide
linkage. The sequence specificity of the enzyme responsible for this modification,
myristoyl CoA:protein N-myristoyl transferase (NMT), has been derived from the
sequence of known N-myristoylated proteins and from studies using synthetic
peptides. The specificity seems to be the following: i.) The N-terminal residue must
20   be glycine; ii.) In position 2, uncharged residues are allowed; iii.) Charged residues,
proline and large hydrophobic residues are not allowed; iv.) In positions 3 and 4,
most, if not all, residues are allowed; v.) In position 5, small uncharged residues are
allowed (Ala, Ser, Thr, Cys, Asn and Gly). Serine is favored; and vi.) In position 6,
proline is not allowed.

25   A consensus pattern for N-myristoylation is as follows: G-{EDRKHPFYW}-
x(2)-[STAGCN]-{P}, wherein 'x' represents any amino acid, and G is the N-
myristoylation site.

         Additional information specific to N-myristoylation sites may be found in
reference to the following publication: Towler D.A., Gordon J.I., Adams S.P., Glaser

L., Annu. Rev. Biochem. 57:69-99(1988); and Grand R.J.A., Biochem. J. 258:625-638(1989); which is hereby incorporated herein in its entirety.

In preferred embodiments, the following N-myristoylation site polypeptides are encompassed by the present invention: ISLAHGIIRSTVLVIF (SEQ ID NO:94),

5   CSMIWGASPSYTILSV (SEQ ID NO:95), MEAKDGSLKAKEGSTG (SEQ ID NO:96), LKAKEGSTGTSESSVE (SEQ ID NO:97), KEGSTGTSESSVEARG (SEQ ID NO:98), TVASDGSMEGKEGSTK (SEQ ID NO:99), HPDLPGTEGGTEGKIV (SEQ ID NO:100), and/or LPGTEGGTEGKIVPSY (SEQ ID NO:101). The present invention also encompasses the use of these N-myristoylation site polypeptides as

10   immunogenic and/or antigenic epitopes as described elsewhere herein.

Moreover, in confirmation of HGPRBMY8 representing a novel GPCR, the HGPRBMY8 polypeptide was predicted to comprise a G-protein coupled receptor motif using the Motif algorithm (Genetics Computer Group, Inc.). G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters,

15   odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. Some examples of receptors that belong to this family are provided as follows: 5-hydroxytryptamine (serotonin) 1A to 1F, 2A to 2C, 4, 5A, 5B, 6 and 7, Acetylcholine, muscarinic-type, M1 to M5, Adenosine A1, A2A, A2B and A3, Adrenergic alpha-1A to -1C; alpha-2A to -2D; beta-1 to –3,

20   Angiotensin II types I and II, Bombesin subtypes 3 and 4, Bradykinin B1 and B2, c3a and C5a anaphylatoxin, Cannabinoid CB1 and CB2, Chemokines C-C CC-CKR-1 to CC-CKR-8, Chemokines C-X-C CXC-CKR-1 to CXC-CKR-4, Cholecystokinin-A and cholecystokinin-B/gastrin, Dopamine D1 to D5, Endothelin ET-a and ET-b, fMet-Leu-Phe (fMLP) (N-formyl peptide), Follicle stimulating hormone (FSH-R), Galanin,

25   Gastrin-releasing peptide (GRP-R), Gonadotropin-releasing hormone (GNRH-R), Histamine H1 and H2 (gastric receptor I), Lutropin-choriogonadotropic hormone (LSH-R), Melanocortin MC1R to MC5R, Melatonin, Neuromedin B (NMB-R), Neuromedin K (NK-3R), Neuropeptide Y types 1 to 6, Neurotensin (NT-R), Octopamine (tyramine) from insects, Odorants, Opioids delta-, kappa- and mu-types,

30   Oxytocin (OT-R), Platelet activating factor (PAF-R), Prostacyclin, Prostaglandin D2,

Prostaglandin E2, EP1 to EP4 subtypes, Prostaglandin F2, Purinoreceptors (ATP),

Somatostatin types 1 to 5, Substance-K (NK-2R), Substance-P (NK-1R), Thrombin,

Thromboxane A2, Thyrotropin (TSH-R), Thyrotropin releasing factor (TRH-R),

Vasopressin V1a, V1b and V2, Visual pigments (opsins and rhodopsin), Proto-

5   oncogene mas, Caenorhabditis elegans putative receptors C06G4.5, C38C10.1,

C43C3.2,T27D1.3 and ZC84.4, Three putative receptors encoded in the genome of

cytomegalovirus: US27, US28, and UL33., ECRF3, a putative receptor encoded in the

genome of herpesvirus saimiri.

The structure of all GPCRs are thought to be identical. They have seven

10   hydrophobic regions, each of which most probably spans the membrane. The N-

terminus is located on the extracellular side of the membrane and is often

glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated.

Three extracellular loops alternate with three intracellular loops to link the seven

transmembrane regions. Most, but not all of these receptors, lack a signal peptide. The

15   most conserved parts of these proteins are the transmembrane regions and the first

two cytoplasmic loops. A conserved acidic-Arg-aromatic triplet is present in the N-

terminal extremity of the second cytoplasmic loop and could be implicated in the

interaction with G proteins.

The putative consensus sequence for GPCRs comprises the conserved triplet

20   and also spans the major part of the third transmembrane helix, and is as follows:

[GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)-

[LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)-[LIVM],

where "X" represents any amino acid.

Additional information relating to G-protein coupled receptors may be found

25   in reference to the following publications: Strosberg A.D., Eur. J. Biochem. 196:1-

10(1991); Kerlavage A.R., Curr. Opin. Struct. Biol. 1:394-401(1991); Probst W.C.,

Snyder L.A., Schuster D.I., Brosius J., Sealfon S.C., DNA Cell Biol. 11:1-20(1992);

Savarese T.M., Fraser C.M., Biochem. J. 283:1-9(1992); Branchek T., Curr. Biol.

3:315-317(1993); Stiles G.L., J. Biol. Chem. 267:6451-6454(1992); Friell T., Kobilka

30   B.K., Lefkowitz R.J., Caron M.G., Trends Neurosci. 11:321-324(1988); Stevens C.F.,

Curr. Biol. 1:20-22(1991); Sakurai T., Yanagisawa M., Masaki T., Trends Pharmacol.

Sci. 13:103-107(1992); Salesse R., Remy J.J., Levin J.M., Jallal B., Garnier J.,

Biochimie 73:109-120(1991); Lancet D., Ben-Arie N., Curr. Biol. 3:668-674(1993);

Uhl G.R., Childers S., Pasternak G., Trends Neurosci. 17:89-93(1994); Barnard E.A.,

5   Burnstock G., Webb T.E., Trends Pharmacol. Sci. 15:67-70(1994); Applebury M.L.,

Hargrave P.A., Vision Res. 26:1881-1895(1986); Attwood T.K., Eliopoulos E.E.,

Findlay J.B.C., Gene 98:153-159(1991); http://www.gcrdb.uthscsa.edu/; and

http://swift.embl-heidelberg.de/7tm/.

In preferred embodiments, the following G-protein coupled receptors

10   signature polypeptide is encompassed by the present invention:

SVVSFIVIPLIVMIACYSVVF (SEQ ID NO:102). Polynucleotides encoding this

polypeptide are also provided. The present invention also encompasses the use of the

HGPRBMY8 G-protein coupled receptors signature polypeptide as immunogenic

and/or antigenic epitopes as described elsewhere herein.

15   For the production of antibodies, various hosts including goats, rabbits,

sheep, rats, mice, humans, and others, can be immunized by injection with

HGPRBMY8 polypeptide, or any fragment or oligopeptide thereof, which has

immunogenic properties. Depending on the host species, various adjuvants may be

used to increase the immunological response. Non-limiting examples of suitable

20   adjuvants include Freund's (complete and incomplete), mineral gels such as

aluminum hydroxide or silica, and surface active substances such as lysolecithin,

pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol.

Adjuvants typically used in humans include BCG (bacilli Calmette Guérin) and

Corynebacterium parvumn.

25   Preferably, the peptides, fragments, or oligopeptides used to induce

antibodies to HGPRBMY8 polypeptide (i.e., immunogens) have an amino acid

sequence having at least five amino acids, and more preferably, at least 7-10 amino

acids. It is also preferable that the immunogens are identical to a portion of the amino

acid sequence of the natural protein; they may also contain the entire amino acid

30   sequence of a small, naturally occurring molecule. The peptides, fragments or

oligopeptides may comprise a single epitope or antigenic determinant or multiple
epitopes. Short stretches of HGPRBMY8 amino acids may be fused with those of
another protein, such as KLH, and antibodies are produced against the chimeric
molecule.

5   Monoclonal antibodies to HGPRBMY8 polypeptide, or immunogenic
fragments thereof, may be prepared using any technique which provides for the
production of antibody molecules by continuous cell lines in culture. These include,
but are not limited to, the hybridoma technique, the human B-cell hybridoma
technique, and the EBV-hybridoma technique (G. Kohler et al., 1975, Nature,

10   256:495-497; D. Kozbor et al., 1985, J. Immunol. Methods, 81:31-42; R.J. Cote et al.,
1983, Proc. Natl. Acad. Sci. USA, 80:2026-2030; and S.P. Cole et al., 1984, Mol. Cell
Biol., 62:109-120). The production of monoclonal antibodies is well known and
routinely used in the art.

           In addition, techniques developed for the production of "chimeric

15   antibodies," the splicing of mouse antibody genes to human antibody genes to obtain
a molecule with appropriate antigen specificity and biological activity can be used
(S.L. Morrison et al., 1984, Proc. Natl. Acad. Sci. USA, 81:6851-6855; M.S.
Neuberger et al., 1984, Nature, 312:604-608; and S. Takeda et al., 1985, Nature,
314:452-454). Alternatively, techniques described for the production of single chain

20   antibodies may be adapted, using methods known in the art, to produce HGPRBMY8
polypeptide-specific single chain antibodies. Antibodies with related specificity, but
of distinct idiotypic composition, may be generated by chain shuffling from random
combinatorial immunoglobulin libraries (D.R. Burton, 1991, Proc. Natl. Acad. Sci.
USA, 88:11120-3). Antibodies may also be produced by inducing *in vivo* production

25   in the lymphocyte population or by screening recombinant immunoglobulin libraries
or panels of highly specific binding reagents as disclosed in the literature (R. Orlandi
et al., 1989, Proc. Natl. Acad. Sci. USA, 86:3833-3837 and G. Winter et al., 1991,
Nature, 349:293-299).

           Antibody fragments, which contain specific binding sites for

30   HGPRBMY8 polypeptide, may also be generated. For example, such fragments
include, but are not limited to, F(ab')₂ fragments which can be produced by pepsin
digestion of the antibody molecule and Fab fragments which can be generated by

reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression

libraries may be constructed to allow rapid and easy identification of monoclonal Fab

fragments with the desired specificity (W.D. Huse et al., 1989, _Science_, 254.1275-

1281).

5   Various immunoassays can be used for screening to identify antibodies

having the desired specificity. Numerous protocols for competitive binding or

immunoradiometric assays using either polyclonal or monoclonal antibodies with

established specificities are well known in the art. Such immunoassays typically

involve measuring the formation of complexes between HGPRBMY8 polypeptide

10   and its specific antibody. A two-site, monoclonal-based immunoassay utilizing

monoclonal antibodies reactive with two non-interfering HGPRBMY8 polypeptide

epitopes is preferred, but a competitive binding assay may also be employed

(Maddox, _supra_).

          Another aspect of the invention relates to a method for inducing an

15   immunological response in a mammal which comprises inoculating the mammal with

HGPRBMY8 polypeptide, or a fragment thereof, adequate to produce antibody and/or

T cell immune response to protect said animal from infections such as bacterial,

fungal, protozoan and viral infections, particularly infections caused by HIV-1 or

HIV-2. Yet another aspect of the invention relates to a method of inducing

20   immunological response in a mammal which comprises, delivering HGPRBMY8

polypeptide via a vector directing expression of HGPRBMY8 polynucleotide _in vivo_

in order to induce such an immunological response to produce antibody to protect said

animal from diseases.

          A further aspect of the invention relates to an immunological/ vaccine

25   formulation (composition) which, when introduced into a mammalian host, induces an

immunological response in that mammal to an HGPRBMY8 polypeptide wherein the

composition comprises an HGPRBMY8 polypeptide or HGPRBMY8 gene. The

vaccine formulation may further comprise a suitable carrier. Since the HGPRBMY8

polypeptide may be broken down in the stomach, it is preferably administered

30   parenterally (including subcutaneous, intramuscular, intravenous, intradermal, etc.,

injection). Formulations suitable for parenteral administration include aqueous and

non-aqueous sterile injection solutions which may contain anti-oxidants, buffers,

bacteriostats and solutes which render the formulation isotonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials, and may be

5   stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in-water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

10   In an embodiment of the present invention, the polynucleotide encoding the HGPRBMY8 polypeptide, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, antisense, to the polynucleotide encoding the HGPRBMY8 polypeptide, may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be

15   transformed with sequences complementary to polynucleotides encoding HGPRBMY8 polypeptide. Thus, complementary molecules may be used to modulate HGPRBMY8 polynucleotide and polypeptide activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers or oligonucleotides, or larger fragments, can be designed from various

20   locations along the coding or control regions of polynucleotide sequences encoding HGPRBMY8 polypeptide.

             Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which

25   are well known to those skilled in the art, can be used to construct recombinant ·vectors which will express a nucleic acid sequence that is complementary to the nucleic acid sequence encoding the HGPRBMY8 polypeptide. These techniques are described both in J. Sambrook et al., *supra* and in F.M. Ausubel et al., *supra*.

             Polypeptides used in treatment can also be generated endogenously in

30   the subject, in treatment modalities often referred to as "gene therapy". Thus for example, cells from a subject may be engineered with a polynucleotide, such as DNA

or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells can then be introduced into the subject.

The genes encoding the HGPRBMY8 polypeptide can be turned off by transforming a cell or tissue with an expression vector that expresses high levels of an

5   HGPRBMY8 polypeptide-encoding polynucleotide, or a fragment thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and even

10   longer if appropriate replication elements are designed to be part of the vector system.

Modifications of gene expression can be obtained by designing antisense molecules or complementary nucleic acid sequences (DNA, RNA, or PNA), to the control, 5', or regulatory regions of the gene encoding the HGPRBMY8 polypeptide, (e.g., signal sequence, promoters, enhancers, and introns).

15   Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent

20   therapeutic advances using triplex DNA have been described (see, for example, J.E. Gee et al., 1994, In: B.E. Huber and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The antisense molecule or complementary sequence may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

25   Ribozymes, i.e., enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Suitable examples include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze

30   endonucleolytic cleavage of sequences encoding HGPRBMY8 polypeptide.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which

64

include the following sequences: GUA, GUU, and GUC. Once identified, short RNA
sequences of between 15 and 20 ribonucleotides corresponding to the region of the
target gene containing the cleavage site may be evaluated for secondary structural
features which may render the oligonucleotide inoperable. The suitability of

5   candidate targets may also be evaluated by testing accessibility to hybridization with
complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes according
to the invention may be prepared by any method known in the art for the synthesis of
nucleic acid molecules. Such methods include techniques for chemically synthesizing

10 oligonucleotides, for example, solid phase phosphoramidite chemical synthesis.
Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription
of DNA sequences encoding HGPRBMY8. Such DNA sequences may be
incorporated into a wide variety of vectors with suitable RNA polymerase promoters
such as T7 or SP. Alternatively, the cDNA constructs that constitutively or inducibly

15 synthesize complementary RNA can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and
half-life. Possible modifications include, but are not limited to, the addition of
flanking sequences at the 5' and/ or 3' ends of the molecule, or the use of
phosphorothioate or 2' O-methyl, rather than phosphodiesterase linkages within the

20 backbone of the molecule. This concept is inherent in the production of PNAs and
can be extended in all of these molecules by the inclusion of nontraditional bases such
as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly
modified forms of adenine, cytosine, guanine, thymine, and uridine which are not as
easily recognized by endogenous endonucleases.

25 Many methods for introducing vectors into cells or tissues are available
and are equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy,
vectors may be introduced into stem cells taken from the patient and clonally
propagated for autologous transplant back into that same patient. Delivery by
transfection and by liposome injections may be achieved using methods, which are

30 well known in the art.

65

Any of the therapeutic methods described above may be applied to any individual in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

A further embodiment of the present invention embraces the

5 administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, diluent, or excipient, for any of the above-described therapeutic uses and effects. Such pharmaceutical compositions may comprise HGPRBMY8 nucleic acid, polypeptide, or peptides, antibodies to HGPRBMY8 polypeptide, mimetics, agonists, antagonists, or inhibitors of

10 HGPRBMY8 polypeptide or polynucleotide. The compositions may be administered alone, or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other

15 agents, drugs, hormones, or biological response modifiers.

The pharmaceutical compositions for use in the present invention can be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual,

20 vaginal, or rectal means.

In addition to the active ingredients (i.e., the HGPRBMY8 nucleic acid or polypeptide, or functional fragments thereof), the pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers or excipients comprising auxiliaries which facilitate processing of the active compounds into preparations

25 which can be used pharmaceutically. Further details on techniques for formulation and administration are provided in the latest edition of <u>Remington's Pharmaceutical Sciences</u> (Mack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable

30 for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained by the combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or

5   protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropyl-methylcellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth, and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as cross-linked polyvinyl

10 pyrrolidone, agar, alginic acid, or a physiologically acceptable salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with physiologically suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide,

15 lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification, or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations, which can be used orally, include push-fit capsules made of gelatin, as well as soft, scaled capsules made of gelatin and a

20 coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

25 Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. In

30 addition, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyloleate or triglycerides, or

liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

5   For topical or nasal administration, penetrants or permeation agents that are appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying,

10   encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, and the like. Salts tend to be more soluble in aqueous solvents, or other protonic solvents, than are the corresponding free base forms. In

15   other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, combined with a buffer prior to use. After the pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For

20   administration of HGPRBMY8 product, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose or

25   amount is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., using neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used and extrapolated to

30   determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example, HGPRBMY8 polypeptide, or fragments thereof, antibodies to

68

HGPRBMY8 polypeptide, agonists, antagonists or inhibitors of HGPRBMY8
polypeptide, which ameliorates, reduces, or eliminates the symptoms or condition.
Therapeutic efficacy and toxicity may be determined by standard pharmaceutical
procedures in cell cultures or experimental animals, e.g., ED_{50} (the dose

5　　therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50%
of the population). The dose ratio of toxic to therapeutic effects is the therapeutic
index, which can be expressed as the ratio, LD_{50}/ED_{50}. Pharmaceutical compositions
which exhibit large therapeutic indices are preferred. The data obtained from cell
culture assays and animal studies are used in determining a range of dosages for

10　　human use. Preferred dosage contained in a pharmaceutical composition is within a
range of circulating concentrations that include the ED_{50} with little or no toxicity.
The dosage varies within this range depending upon the dosage form employed,
sensitivity of the patient, and the route of administration.

　　　　　　The practitioner, who will consider the factors related to the individual

15　　requiring treatment, will determine the exact dosage. Dosage and administration are
adjusted to provide sufficient levels of the active moiety or to maintain the desired
effect. Factors, which may be taken into account, include the severity of the
individual's disease state, general health of the patient, age, weight, and gender of the
patient, diet, time and frequency of administration, drug combination(s), reaction

20　　sensitivities, and tolerance/ response to therapy. As a general guide, long-acting
pharmaceutical compositions may be administered every 3 to 4 days, every week, or
once every two weeks, depending on half-life and clearance rate of the particular
formulation. Variations in these dosage levels can be adjusted using standard
empirical routines for optimization, as is well understood in the art.

25　　　　　　Normal dosage amounts may vary from 0.1 to 100,000 micrograms
(μg), up to a total dose of about 1 gram (g), depending upon the route of
administration. Guidance as to particular dosages and methods of delivery is
provided in the literature and is generally available to practitioners in the art. Those
skilled in the art will employ different formulations for nucleotides than for proteins

30　　or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be
specific to particular cells, conditions, locations, and the like.

In another embodiment of the present invention, antibodies which
specifically bind to the HGPRBMY8 polypeptide may be used for the diagnosis of
conditions or diseases characterized by expression (or overexpression) of the
HGPRBMY8 polynucleotide or polypeptide, or in assays to monitor patients being

5   treated with the HGPRBMY8 polypeptide, or its agonists, antagonists, or inhibitors.
The antibodies useful for diagnostic purposes may be prepared in the same manner as
those described above for use in therapeutic methods. Diagnostic assays for the
HGPRBMY8 polypeptide include methods, which utilize the antibody and a label to
detect the protein in human body fluids or extracts of cells or tissues. The antibodies

10   may be used with or without modification, and may be labeled by joining them, either
covalently or non-covalently, with a reporter molecule. A wide variety of reporter
molecules, which are known in the art, may be used, several of which are described
above. . In particular, a method of detecting a G-protein coupled receptor,
homologue, or an antibody-reactive fragment thereof, in a sample, comprising: a)

15   contacting the sample with an antibody specific for the polypeptide, or an antigenic
fragment thereof, under conditions in which an antigen-antibody complex can form
between the antibody and the polypeptide or antigenic fragment thereof in the sample;
and b) detecting an antigen-antibody complex formed in step (a), wherein detection
of the complex indicates the presence of an antigenic fragment thereof, in the sample.

20   The use of mammalian cell reporter assays to demonstrate functional
coupling of known GPCRs (G Protein Coupled Receptors) has been well documented
in the literature (Gilman, 1987, Boss et al., 1996; Alam & Cook, 1990; George et al.,
1997; Selbie & Hill, 1998; Rees et al., 1999). In fact, reporter assays have been
successfully used for identifying novel small molecule agonists or antagonists against

25   GPCRs as a class of drug targets (Zlokarnik et al., 1998; George et al., 1997; Boss et
al., 1996; Rees et al, 2001). In such reporter assays, a promoter is regulated as a
direct consequence of activation of specific signal transduction cascades following
agonist binding to a GPCR (Alam & Cook 1990; Selbie & Hill, 1998; Boss et al.,
1996; George et al., 1997; Gilman, 1987).

30   A number of response element-based reporter systems have been
developed that enable the study of GPCR function. These include cAMP response
element (CRE)-based reporter genes for G alpha i/o, G alpha s- coupled GPCRs,

70

Nuclear Factor Activator of Transcription (NFAT)-based reporters for G alpha q/11

11or the promiscuous G protein G alpha 15/16 –coupled receptors and MAP kinase

reporter genes for use in Galpha i/o coupled receptors (Selbie & Hill, 1998; Boss et

al., 1996; George et al., 1997; Blahos, et al., 2001; Offermann & Simon, 1995;

5   Gilman, 1987; Rees et al., 2001). Transcriptional response elements that regulate the

expression of Beta-Lactamase within a CHO K1 cell line (CHO-NFAT/CRE: Aurora

Biosciences ™) (Zlokarnik et al., 1998) have been implemented to characterize the

function of the orphan HGPRBMY8 polypeptide of the present invention. The

system enables demonstration of constitutive G-protein coupling to endogenous

10   cellular signaling components upon intracellular overexpression of orphan receptors.

Overexpression has been shown to represent a physiologically relevant event. For

example, it has been shown that overexpression occurs in nature during metastatic

carcinomas, wherein defective expression of the monocyte chemotactic protein 1

receptor, CCF2, in macrophages is associated with the incidence of human ovarian

15   carcinoma (Sica, et al.,2000; Salcedo et al., 2000). Indeed, it has been shown that

overproduction of the Beta 2 Adrenergic Receptor in transgenic mice leads to

constitutive activation of the receptor signaling pathway such that these mice exhibit

increased cardiac output (Kypson et al., 1999; Dorn et al., 1999). These are only a

few of the many examples demonstrating constitutive activation of GPCRs whereby

20   many of these receptors are likely to be in the active, R*, conformation (J.Wess 1997)

(see Example 7).

         Several assay protocols including ELISA, RIA, and FACS for

measuring HGPRBMY8 polypeptide are known in the art and provide a basis for

diagnosing altered or abnormal levels of HGPRBMY8 polypeptide expression.

25   Normal or standard values for HGPRBMY8 polypeptide expression are established by

combining body fluids or cell extracts taken from normal mammalian subjects,

preferably human, with antibody to the HGPRBMY8 polypeptide under conditions

suitable for complex formation. The amount of standard complex formation may be

quantified by various methods; photometric means are preferred. Quantities of

30   HGPRBMY8 polypeptide expressed in subject sample, control sample, and disease

samples from biopsied tissues are compared with the standard values. Deviation

between standard and subject values establishes the parameters for diagnosing disease.

Microarrays and Screening Assays

5   In another embodiment of the present invention, oligonucleotides, or longer fragments derived from the HGPRBMY8 polynucleotide sequence described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This

10   information may be used to determine gene function, to understand the genetic basis of a disease, to diagnose disease, and to develop and monitor the activities of therapeutic agents. In a particular aspect, the microarray is prepared and used according to the methods described in WO 95/11995 (Chee et al.); D.J. Lockhart et al., 1996, Nature Biotechnology, 14:1675-1680; and M. Schena et al., 1996, Proc.

15   Natl. Acad. Sci. USA, 93:10614-10619). Microarrays are further described in U.S. Patent No. 6,015,702 to P. Lal et al.

In another embodiment of this invention, the nucleic acid sequence, which encodes the HGPRBMY8 polypeptide, may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic

20   sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial PI · constructions, or single chromosome cDNA libraries, as reviewed by C.M. Price, 1993, *Blood Rev.*, 7:127-134 and by B.J. Trask, 1991, *Trends Genet.*, 7:149-154.

25   Fluorescent *In Situ* Hybridization (FISH), (as described in I. Verma et al., 1988, Human Chromosomes: A Manual of Basic Techniques Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in numerous scientific journals, or at Online Mendelian Inheritance in Man (OMIM).

30   Correlation between the location of the gene encoding the HGPRBMY8 polypeptide on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease.

72

The nucleotide sequences, particularly that of SEQ ID NO:1, or fragments thereof, according to this invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical

5   mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers, even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical

10   mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (R.A. Gatti et al., 1988, Nature, 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for

15   further investigation. The nucleotide sequence of the present invention may also be used to detect differences in the chromosomal location due to translocation, inversion, and the like, among normal, carrier, or affected individuals.

[0100] In another embodiment of the present invention, the HGPRBMY8 polypeptide, its catalytic or immunogenic fragments or oligopeptides thereof, can be

20   used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between HGPRBMY8 polypeptide, or portion thereof, and the agent being tested, may be measured utilizing techniques commonly

25   practiced in the art. In particular, a method of screening a library of molecules or compounds with an HGPRBMY8 polynucleotide, or fragment thereof, to identify at least one molecule or compound therein which specifically binds to the G-protein coupled receptor polynucleotide sequence, preferably the HGPRBMY8 polynucleotide sequence, or fragment thereof, comprising: a) combining the G-

30   protein coupled receptor polynucleotide, or fragment thereof, with a library of molecules or compounds under conditions to allow specific binding; and b) detecting specific binding, thereby identifying a molecule or compound, which specifically

binds to a G-protein coupled receptor-encoding polynucleotide sequence. In a further embodiment, the screening method is a high throughput screening method. Preferably, the library is selected from the group consisting of DNA molecules, RNA molecules, artificial chromosome constructions, PNAs, peptides and proteins. In

5   another preferred embodiment, the candidate small molecules or compounds are a drug or therapeutic.

In yet another embodiment, a method of screening for candidate compounds capable of modulating activity of a G-protein coupled receptor-encoding polypeptide, comprising: a) contacting a test compound with a cell or tissue

10   expressing the G-protein coupled receptor polypeptide, homologue, or fragment thereof; and b) selecting as candidate modulating compounds those test compounds that modulate activity of the G-protein coupled receptor polypeptide. Preferably, the candidate compounds are agonists or antagonists of G-protein coupled receptor activity. More preferably, the polypeptide activity is associated with the brain.

15   Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in WO 84/03564 (Venton, et al.). In this method, as applied to the HGPRBMY8 protein, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The

20   test compounds are reacted with the HGPRBMY8 polypeptide, or fragments thereof, and washed. Bound HGPRBMY8 polypeptide is then detected by methods well known in the art. Purified HGPRBMY8 polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid

25   support.

In a further embodiment of this invention, competitive drug screening assays can be used in which neutralizing antibodies, capable of binding the HGPRBMY8 polypeptide, specifically compete with a test compound for binding to the HGPRBMY8 polypeptide. In this manner, the antibodies can be used to detect the

30   presence of any peptide, which shares one or more antigenic determinants with the HGPRBMY8 polypeptide.

Other screening and small molecule (e.g., drug) detection assays which involve the detection or identification of small molecules or compounds that can bind to a given protein, i.e., the HGPRBMY8 polypeptide, are encompassed by the present invention. Particularly preferred are assays suitable for high throughput screening

5　　methodologies. In such binding-based screening or detection assays, a functional assay is not typically required. All that is needed is a target protein, preferably substantially purified, and a library or panel of compounds (e.g., ligands, drugs, small molecules) to be screened or assayed for binding to the protein target. Preferably, most small molecules that bind to the target protein will modulate activity in some

10　　manner, due to preferential, higher affinity binding to functional areas or sites on the protein.

An example of such an assay is the fluorescence based thermal shift assay (3-Dimensional Pharmaceuticals, Inc., 3DP; Exton, PA) as described in U.S. Patent Nos. 6,020,141 and 6,036,920 to Pantoliano et al.; see also, J. Zimmerman,

15　　2000, *Gen. Eng. News*, 20(8)). The assay allows the detection of small molecules (e.g., drugs, ligands) that bind to expressed, and preferably purified, HGPRBMY8 polypeptide based on affinity of binding determinations by analyzing thermal unfolding curves of protein-drug or ligand complexes. The drugs or binding molecules determined by this technique can be further assayed, if desired, by

20　　methods, such as those described herein, to determine if the molecules affect or modulate function or activity of the target protein.

EXAMPLES

[0101]　　The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in

25　　any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the construction of vectors, the insertion of cDNA into such vectors, or the introduction of the resulting vectors into the appropriate host. Such methods are well known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, Molecular

30　　Cloning: a Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1
BIOINFORMATICS ANALYSIS

G-protein coupled receptor sequences were used as a probes to search
human genomic sequence databases. The search program used was gapped BLAST

5   (S.F. Altschul, et al., Nuc. Acids Res., 25:3389-4302 (1997)). The top genomic exon
hits from the BLAST results were searched back against the non-redundant protein
and patent sequence databases. From this analysis, exons encoding potential full-
length sequence of a novel human GPCR, HGPRBMY8, was identified directly from
the genomic sequence. The full-length clone of this GPCR was experimentally

10   obtained by RT-PCR using the sequence from genomic data. The complete protein
sequence of HGPRBMY8 was analyzed for potential transmembrane domains.
TMPRED program (K. Hofmann and W. Stoffel, Biol. Chem., 347:166 (1993) was
used for transmembrane prediction. The program predicted seven transmembrane
domains and the predicted domains match with the predicted transmembrane domains

15   of related GPCRs at the sequence level. Based on sequence, structure and known
GPCR signature sequences, the orphan protein, HGPRBMY8 of the present invention,
is a novel human GPCR.

EXAMPLE 2
CLONING OF THE NOVEL HUMAN GPCR HGPRBMY8

20   HGPRBMY8 was cloned from a human brain cDNA library (Clontech;
Palo Alto, CA) by PCR amplification of the predicted cDNA sequence using
sequence specific oligonucleotides. The 5' sense oligonucleotide was as follows:

            5'-GGCCGAATTCGCAACCTGTCTCACGCCCTCTGG-3'

(SEQ ID NO:5). The 3' anti-sense oligonucleotide was as follows:

25   5'-
GGCCGAATTCGGACAGTTCAAGGTTTGCCTTAGAAC-3' (SEQ ID NO:6).
These oligonucleotides contained EcoRI restriction enzyme sites for subcloning the
PCR fragment into the mammalian expression vector, pcDNA6. Samples containing
human brain cDNA, the 5 prime sense, and 3 prime anti-sense oligonucleotides were

30   subjected to PCR amplification followed by gel purification of the amplified product.
The inserts of cDNA clones that were positive by PCR were sized, and two of the
largest clones (~1.6 Kb) were sequenced using conventional sequencing methods.

Purified sample was digested with EcoRI, extracted with phenol:chloroform, and ligated into pcDNA6. The resultant plasmids were subjected to DNA sequencing and the sequences were verified by comparison with the database sample.

EXAMPLE 3

5   EXPRESSION PROFILING OF NOVEL HUMAN GPCR, HGPRBMY8

The oligonucleotides used for the expression profiling of HGPRBMY8 are:

HGPRBMY8-2s: 5'-GCAGAGCACTCCTCCACTCT-3'

(SEQ ID NO:34)

10   HGPRBMY8-2a: 5'-AGCAGGCAATCATGACAATC-3'

(SEQ ID NO:35)

These oligonucleotides were used to measure the steady state levels of mRNA by quantitative PCR. Briefly, first strand cDNA was made from commercially available mRNA (Clontech; Palo Alto, CA). The relative amount of

15   cDNA used in each assay (2.5 ng of cDNA per assay) was determined by performing a parallel experiment using a primer pair for the cyclophilin gene, which is expressed in equal amounts in all tissues. The cyclophilin primer pair detected small variations in the amount of cDNA in each sample, and these data were used for normalization of the data obtained with the primer pair for HGPRBMY8. The PCR data were

20   converted into a relative assessment of the difference in transcript abundance among the tissues tested and the data are presented in Figure 7. Transcripts corresponding to the orphan GPCR, HGPRBMY8, were found to be highly expressed in brain.

EXAMPLE 4

G-PROTEIN COUPLED RECEPTOR PCR EXPRESSION PROFILING

25   Based on HGPRBMY8's expression in the brain, further analysis was carried out to determine if there was any additional specificity within sub regions. The same PCR primer pair that was used to identify HGPRBMY8 (also referred to as GPCR 58 and GPCR84) cDNA clones was used to measure the steady state levels of mRNA by quantitative PCR.

30

GPCR84-s   GTTAGCCTCACCCACCTGTT   (SEQ ID NO:36)

77

GPCR84-a   CACAATCCAGGTGCCATAGA   (SEQ ID NO:37)

Briefly, first strand cDNA was made from commercially available brain subregion mRNA (Clontech) and subjected to real time quantitative PCR using a PE 5700 instrument (Applied Biosystems; Foster City, CA) which detects the

5   amount of DNA amplified during each cycle by the fluorescent output of SYBR green, a DNA binding dye specific for double strands. The specificity of the primer pair for its target is verified by performing a thermal denaturation profile at the end of the run which gives an indication of the number of different DNA sequences present by determining melting Tm. In the case of the HGPRBMY8 primer pair, only one

10   DNA fragment was detected having a homogeneous melting point. Contributions of contaminating genomic DNA to the assessment of tissue abundance is controlled for by performing the PCR with first strand made with and without reverse transcriptase. In all cases, the contribution of material amplified in the no reverse transcriptase controls was negligible.

15   More specifically, since HGPRBMY8 is expressed at extremely low levels, each PCR reaction contained the amount of first strand cDNA made from 100 nanograms of poly A+ RNA (2.5 nanograms is the standard amount).

The number of reactions and amount of mix needed was first determined. All of the samples were run in triplicate, so sample tubes needed 3.5

20   reactions worth of mixture using the following formula as a guide (2x # tissue samples + 1 no template control + 1 for pipetting error)(3.5).

The reaction mixture consisted of the following components and volumes:

COMPONENTS	VOL/RXN
2X SybrGreen Master Mix	25 microliters
water	23.5 microliters
primer mix (10uM ea.)	0.5 microliters
cDNA (100ng/uL)	1 microliter

25   The mixture was initially made without cDNA for enough reactions as determined

78

above. The mix (171.5 μl) was then aliquoted into sample tubes. cDNA (3.5 μl) was added to each sample tube, mixed gently, and spun down for collection. Three 50 μl samples were aliquoted to the optical plate, where the primer and sample were set up for sample analysis. The threshold was set in Log view to intersect linear regions of

5   amplification. The background was set in Linear view to 2-3 cycles before the amplification curve appears. The mean values for RT+ was calculated and normalized to Cyclophilin: dc_t = sample mean – cyclophilin mean. The ddc_t was determined by subtracting individual dcts from the highest value of dc_t in the list. The relative abundance was determined by formula $2^{\wedge}ddc_t$.

10   Small variations in the amount of cDNA used in each tube was determined by performing a parallel experiment using a primer pair for a gene expressed in equal amounts in all tissues, cyclophilin. These data were used to normalize the data obtained with the HGPRBMY8 primer pair. The PCR data was converted into a relative assessment of the difference in transcript abundance amongst

15   the tissues tested and the data are presented in bar graph form. Transcripts corresponding to HGPRBMY8 are expressed approximately 825 times greater in the caudate nucleus than in the substantia nigra. Low level expression was detected in the thalamus, amygdala, hippocampus, cerebellum and corpus collosum (see FIG. 8).

EXAMPLE 5
20   SIGNAL TRANSDUCTION ASSAYS
            The activity of GPCRs or homologues thereof, can be measured using any assay suitable for the measurement of the activity of a G protein-coupled receptor, as commonly known in the art. Signal transduction activity of a G protein-coupled receptor can be monitor by monitoring intracellular Ca^{2+}, cAMP, inositol 1,4,5-

25   triphosphate (IP_3), or 1,2-diacylglycerol (DAG). Assays for the measurement of intracellular Ca^{2+} are described in Sakurai et al. (EP 480 381). Intracellular IP_3 can be measured using a kit available from Amersham, Inc. (Arlington Heights, IL). A kit for measuring intracellular cAMP is available from Diagnostic Products, Inc. (Los Angeles, CA).

30   Activation of a G protein-coupled receptor triggers the release of Ca^{2+} ions sequestered in the mitochondria, endoplasmic reticulum, and other cytoplasmic

vesicles into the cytoplasm. Fluorescent dyes, e.g., fura-2, can be used to measure the

concentration of free cytoplasmic Ca^{2+}. The ester of fura-2, which is lipophilic and

can diffuse across the cell membrane, is added to the media of the host cells

expressing GPCRs. Once inside the cell, the fura-2 ester is hydrolyzed by cytosolic

5   esterases to its non-lipophilic form, and then the dye cannot diffuse back out of the

cell. The non-lipophilic form of fura-2 will fluoresce when it binds to free Ca^{2+}. The

fluorescence can be measured without lysing the cells at an excitation spectrum of

340 nm or 380 nm and at fluorescence spectrum of 500 nm (Sakurai et al., EP 480

381).

10   Upon activation of a G protein-coupled receptor, the rise of free

cytosolic Ca^{2+} concentrations is preceded by the hydrolysis of phosphatidylinositol

4,5-bisphosphate. Hydrolysis of this phospholipid by the phospholipase C yields 1,2-

diacylglycerol (DAG), which remains in the membrane, and water-soluble inositol

1,4,5-triphosphate (IP_3). Binding of ligands or agonists will increase the

15   concentration of DAG and IP3. Thus, signal transduction activity can be measured by

monitoring the concentration of these hydrolysis products.

To measure the IP_3 concentrations, radioactivity labeled ^{3}H-inositol is

added to the media of host cells expressing GPCRs. The ^{3}H-inositol is taken up by

the cells and incorporated into IP_3. The resulting inositol triphosphate is separated

20   from the mono and di-phosphate forms and measured (Sakurai et al., EP 480 381).

Alternatively, Amersham provides an inositol 1,4,5-triphosphate assay system. With

this system Amersham provides tritylated inositol 1,4,5-triphosphate and a receptor

capable of distinguishing the radioactive inositol from other inositol phosphates.

With these reagents an effective and accurate competition assay can be performed to

25   determine the inositol triphosphate levels.

Cyclic AMP levels can be measured according to the methods

described in Gilman et al., Proc. Natl. Acad. Sci. 67:305-312 (1970). In addition, a

kit for assaying levels of cAMP is available from Diagnostic Products Corp. (Los

Angeles, CA).

EXAMPLE 6
GPCR ACTIVITY

Another method for screening compounds which are antagonists, and thus inhibit activation of the receptor polypeptide of the present invention is provided.

5   This involves determining inhibition of binding of labeled ligand, such as dATP, dAMP, or UTP, to cells which have the receptor on the surface thereof, or cell membranes containing the receptor. Such a method further involves transfecting a eukaryotic cell with DNA encoding the GPCR polypeptide such that the cell expresses the receptor on its surface. The cell is then contacted with a potential

10   antagonist in the presence of a labeled form of a ligand, such as dATP, dAMP, or UTP. The ligand can be labeled, e.g., by radioactivity, fluorescence, or any detectable label commonly known in the art. The amount of labeled ligand bound to the receptors is measured, e.g., by measuring radioactivity associated with transfected cells or membrane from these cells. If the compound binds to the receptor, the

15   binding of labeled ligand to the receptor is inhibited as determined by a reduction of labeled ligand which binds to the receptors. This method is called a binding assay. Naturally, this same technique can be used to determine agonists.

In a further screening procedure, mammalian cells, for example, but not limited to, CHO, HEK 293, Xenopus Oocytes, RBL-2H3, etc., which are

20   transfected, are used to express the receptor of interest. The cells are loaded with an indicator dye that produces a fluorescent signal when bound to calcium, and the cells are contacted with a test substance and a receptor agonist, such as DATP, DAMP, or UTP. Any change in fluorescent signal is measured over a defined period of time using, for example, a fluorescence spectrophotometer or a fluorescence imaging plate

25   reader. A change in the fluorescence signal pattern generated by the ligand indicates that a compound is a potential antagonist or agonist for the receptor.

In yet another screening procedure, mammalian cells are transfected to express the receptor of interest, and are also transfected with a reporter gene construct that is coupled to activation of the receptor (for example, but not limited to luciferase

30   or beta-galactosidase behind an appropriate promoter). The cells are contacted with a test substance and the receptor agonist (ligand), such as dATP, dAMP, or UTP, and the signal produced by the reporter gene is measured after a defined period of time.

81

The signal can be measured using a luminometer, spectrophotometer, fluorimeter, or

other such instrument appropriate for the specific reporter construct used. Inhibition

of the signal generated by the ligand indicates that a compound is a potential

antagonist for the receptor.

5   Another screening technique for antagonists or agonists involves

introducing RNA encoding the GPCR polypeptide into cells (or CHO, HEK 293,

RBL-2H3, etc.) to transiently or stably express the receptor. The receptor cells are

then contacted with the receptor ligand, such as dATP, dAMP, or UTP, and a

compound to be screened. Inhibition or activation of the receptor is then determined

10   by detection of a signal, such as, cAMP, calcium, proton, or other ions.

EXAMPLE 7
FUNCTIONAL CHARACTERIZATION OF HGPRBMY8

          The putative GPCR HGPRBMY8 cDNA was PCR amplified using

PFU™ (Stratagene). The primers used in the PCR reaction were specific to the

15   HGPRBMY8 polynucleotide and were ordered from Gibco BRL (5 prime primer: 5'-

GTCCCCAAGCTTGCACCATGACGTCCACCTGCACCAACAGCA -3' (SEQ ID

NO:38). The following 3 prime primer was used to add a Flag-tag epitope to the

HGPRBMY8 polypeptide for immunocytochemistry: 5'-

CGGGATCCTACTTGTCGTCGTCGTCCTTGTAGTCCATAGGAAAAGTAGCAG

20   AATCGTAGGAA-3' (SEQ ID NO:39). The product from the PCR reaction was

isolated from a 0.8% Agarose gel (Invitrogen) and purified using a Gel Extraction Kit

™ from Qiagen.

          The purified product was then digested overnight along with the

pcDNA3.1 Hygro ™ mammalian expression vector from Invitrogen using the HindIII

25   and BamHI restriction enzymes (New England Biolabs). These digested products

were then purified using the Gel Extraction Kit ™ from Qiagen and subsequently

ligated to the pcDNA3.1 Hygro ™ expression vector using a DNA molar ratio of 4

parts insert: 1 vector. All DNA modification enzymes were purchased from NEB.

The ligation was incubated overnight at 16 degrees Celsius, after which time, one

30   microliter of the mix was used to transform DH5 alpha cloning efficiency competent

E. coli ™ (Gibco BRL). A detailed description of the pcDNA3.1 Hygro ™

82

mammalian expression vector is available at the Invitrogen web site
(www.Invitrogen.com). The plasmid DNA from the ampicillin resistant clones were
isolated using the Wizard DNA Miniprep System ™ from Promega. Positive clones
were then confirmed and scaled up for purification using the Qiagen Maxiprep ™

5   plasmid DNA purification kit.

Cell Line Generation:

The pcDNA3.1hygro vector containing the orphan HGPRBMY8
cDNA were used to transfect CHO-NFAT/CRE (Aurora Biosciences) cells using

10   Lipofectamine 2000 ™ according to the manufacturers specifications (Gibco BRL).
Two days later, the cells were split 1:3 into selective media (DMEM 11056, 600
μg/ml Hygromycin, 200 μg/ml Zeocin, 10% FBS). All cell culture reagents were
purchased from Gibco BRL-Invitrogen.

The CHO-NFAT/CRE and the CHO-NFAT G alpha 15 cell lines,

15   transiently or stably transfected with the orphan HGPRBMY8 GPCR, were analyzed
using the FACS Vantage SE ™ (BD), fluorescence microscopy (Nikon), and the LJL
Analyst ™ (Molecular Devices). In this system, changes in real-time gene
expression, as a consequence of constitutive G-protein coupling of the orphan
HGPRBMY8 GPCR, is examined by analyzing the fluorescence emission of the

20   transformed cells at 447nm and 518nm. The changes in gene expression can be
visualized using Beta-Lactamase as a reporter, that, when induced by the appropriate
signaling cascade, hydrolyzes an intracellularly loaded, membrane-permeant ester
substrate Cephalosporin-Coumarin-Fluorescein2/ Acetoxymethyl (CCF2/AM™
Aurora Biosciences; Zlokarnik, et al., 1998). The CCF2/AM™ substrate is a 7-

25   hydroxycoumarin cephalosporin with a fluorescein attached through a stable thioether
linkage. Induced expression of the Beta-Lactamase enzyme is readily apparent since
each enzyme molecule produced is capable of changing the fluorescence of many
CCF2/AM ™ substrate molecules. A schematic of this cell based system is shown
below.
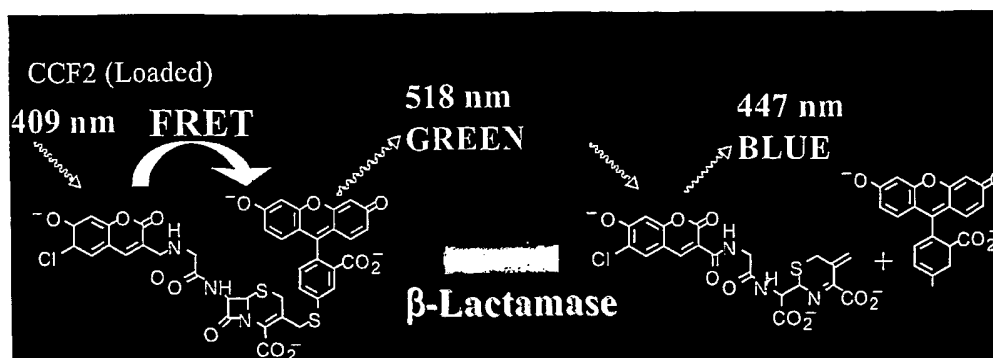
In summary, CCF2/AM ™ is a membrane permeant, intracellularly-
trapped, fluorescent substrate with a cephalosporin core that links a 7-
hydroxycoumarin to a fluorescein. For the intact molecule, excitation of the coumarin
at 409 nm results in Fluorescence Resonance Energy Transfer (FRET) to the
fluorescein which emits green light at 518 nm. Production of active Beta-Lactamase
results in cleavage of the Beta-Lactam ring, leading to disruption of FRET, and
excitation of the coumarin only - thus giving rise to blue fluorescent emission at 447
nm.

Fluorescent emissions were detected using a Nikon-TE300 microscope
equipped with an excitation filter (D405/10X-25), dichroic reflector (430DCLP), and
a barrier filter for dual DAPI/FITC (510nM) to visually capture changes in Beta-
Lactamase expression. The FACS Vantage SE is equiped with a Coherent Enterprise
II Argon Laser and a Coherent 302C Krypton laser. In flow cytometry, UV excitation
at 351-364 nm from the Argon Laser or violet excitation at 407 nm from the Krypton
laser are used. The optical filters on the FACS Vantage SE are HQ460/50m and
HQ535/40m bandpass separated by a 490 dichroic mirror.

Prior to analyzing the fluorescent emissions from the cell lines as
described above, the cells were loaded with the CCF2/AM substrate. A 6X CCF2/AM
loading buffer was prepared whereby 1mM CCF2/AM (Aurora Biosciences) was
dissolved in 100% DMSO (Sigma). Stock solution (12 µl) was added to 60 µl of
100mg/ml Pluronic F127 (Sigma) in DMSO containing 0.1% Acetic Acid (Sigma).
This solution was added while vortexing to 1 mL of Sort Buffer (PBS minus calcium
and magnesium-Gibco-25 mM HEPES-Gibco- pH 7.4, 0.1% BSA). Cells were
placed in serum-free media and the 6X CCF2/AM was added to a final concentration

of 1X. The cells were then loaded at room temperature for one to two hours, and then
subjected to fluorescent emission analysis as described herein. Additional details
relative to the cell loading methods and/or instrument settings may be found by
reference to the following publications: see Zlokarnik, et al., 1998; Whitney et al.,

5   1998; and BD Biosciences,1999.

Immunocytochemistry:

        The cell lines transfected and selected for expression of Flag-epitope
tagged orphan GPCRs were analyzed by immunocytochemistry. The cells were

10   plated at 1X10³ in each well of a glass slide (VWR). The cells were rinsed with PBS
followed by acid fixation for 30 minutes at room temperature using a mixture of 5%
Glacial Acetic Acid / 90% ethanol. The cells were then blocked in 2% BSA and
0.1%Triton in PBS, incubated for 2 h at room temperature or overnight at 4°C. A
monoclonal FITC antibody directed against FLAG was diluted at 1:50 in blocking

15   solution and incubated with the cells for 2 h at room temperature. Cells were then
washed three times with 0.1%Triton in PBS for five minutes. The slides were
overlayed with mounting media dropwise with Biomedia –Gel Mount™ (Biomedia;
Containing Anti-Quenching Agent). Cells were examined at 10x magnification using
the Nikon TE300 equiped with FITC filter (535nm).

20   There is strong evidence that certain GPCRs exhibit a cDNA
concentration-dependent constitutive activity through cAMP response element (CRE)
luciferase reporters (Chen et al., 1999). In an effort to demonstrate functional
coupling of HGPRBMY8 to known GPCR second messenger pathways, the
HGPRBMY8 polypeptide was expressed at high constitutive levels in the CHO-

25   NFAT/CRE cell line. To this end, the HGPRBMY8 cDNA was PCR amplified and
subcloned into the pcDNA3.1 hygro ™ mammalian expression vector as described
herein. Early passage CHO-NFAT/CRE cells were then transfected with the resulting
pcDNA3.1 hygro ™ / HGPRBMY8 construct. Transfected and non-transfected
CHO-NFAT/CRE cells (control) were loaded with the CCF2 substrate and stimulated

30   with 10 nM PMA, 1 µM Thapsigargin (NFAT stimulator), and 10 µM Forskolin (CRE

stimulator) to fully activate the NFAT/CRE element. The cells were then analyzed for fluorescent emission by FACS.

The FACS profile demonstrates the constitutive activity of HGPRBMY8 in the CHO-NFAT/CRE line as evidenced by the significant population

5   of cells with blue fluorescent emission at 447 nm (see Figure 12: Blue Cells). Figure 12 further describes CHO-NFAT/CRE cell lines transfected with the pcDNA3.1 Hygro ™ / HGPRBMY8 mammalian expression vector. The cells were analyzed via FACS according to their wavelength emission at 518 nM (Channel R3 - Green Cells), and 447 nM (Channel R2 – Blue Cells). As shown, overexpression of HGPRBMY8

10   results in functional coupling and subsequent activation of beta lactamase gene expression, as evidenced by the significant number of cells with fluorescent emission at 447 nM relative to the non-transfected control CHO-NFAT/CRE cells (shown in Figure 11). As expected, the NFAT/CRE response element in the untransfected control cell line was not activated (i.e., beta lactamase not induced), enabling the

15   CCF2 substrate to remain intact, and resulting in the green fluorescent emission at 518 nM (see Figure 11-Green Cells). Figure 11 describes control CHO-NFAT/CRE (Nuclear Factor Activator of Transcription (NFAT) / cAMP response element (CRE)) cell lines, in the absence of the pcDNA3.1 Hygro ™ / HGPRBMY8 mammalian expression vector transfection. The cells were analyzed via FACS (Fluorescent

20   Assisted Cell Sorter) according to their wavelength emission at 518 nM (Channel R3 - Green Cells), and 447 nM (Channel R2 – Blue Cells). As shown, the vast majority of cells emit at 518 nM, with minimal emission observed at 447 nM. The latter is expected since the NFAT/CRE response elements remain dormant in the absence of an activated G-protein dependent signal transduction pathway (e.g., pathways

25   mediated by Gq/11 or Gs coupled receptors). As a result, the cell permeant, CCF2/AM™ (Aurora Biosciences; Zlokarnik, et al., 1998) substrate remains intact and emits light at 518 nM.

A very low level of leaky Beta Lactamase expression was detectable as evidenced by the small population of cells emitting at 447 nm. Analysis of a stable

30   pool of cells transfected with HGPRBMY8 revealed constitutive coupling of the cell population to the NFAT/CRE response element, activation of Beta Lactamase and cleavage of the substrate (Figure 12-Blue Cells). These results demonstrate that

86

overexpression of HGPRBMY8 leads to constitutive coupling of signaling pathways known to be mediated by Gq/11or G alpha 15/16 or Gs coupled receptors that converge to activate either the NFAT or CRE response elements respectively (Boss et al., 1996; Chen et al., 1999).

5   In an effort to further characterize the observed functional coupling of the HGPRBMY8 polypeptide, its ability to couple to the cAMP response element (CRE) independent of the NFAT response element was examined. To this end, HEK-CRE cell line that contained only the integrated 3XCRE linked to the Beta-Lactamase reporter was transfected with the pcDNA3.1 hygro ™ / HGPRBMY8 construct.

10   Analysis of the fluorescence emission from this stable pool showed that HGPRBMY8 constitutively coupled to the cAMP mediated second messenger pathways (see Figure 14 relative to Figure 13). Figure 14 describes FACS analysis of HEK-CRE cell lines transfected with the pcDNA3.1 Hygro ™ / HGPRBMY8 mammalian expression vector according to their wavelength emission at 518 nM (Channel R3 - Green Cells),

15   and 447 nM (Channel R2 – Blue Cells). As shown, overexpression of HGPRBMY8 in the HEK-CRE cells resulted in functional coupling, as evidenced by the insignificant background level of cells with fluorescent emission at 447 nM. Figure 13 describes HEK-CRE cell lines in the absence of the pcDNA3.1 Hygro ™ / HGPRBMY8 mammalian expression vector transfection. The cells were analyzed via

20   FACS (Fluorescent Assisted Cell Sorter) according to their wavelength emission at 518 nM (Channel R3 - Green Cells), and 447 nM (Channel R2 – Blue Cells). As shown, the vast majority of cells emit at 518 nM, with minimal emission observed at 447 nM. The latter is expected since the CRE response elements remain dormant in the absence of an activated G-protein dependent signal transduction pathway (e.g.,

25   pathways mediated by Gs coupled receptors). As a result, the cell permeant, CCF2/AM™ (Aurora Biosciences; Zlokarnik, et al., 1998) substrate remains intact and emits light at 518 nM.

Experiments have shown that known G coupled receptors do demonstrate constitutive activation when overexpressed in the HEK-CRE cell line.

30   For example, direct activation of adenylate cyclase using 10 µM Forskolin has been shown to activate CRE and the subsequent induction of Beta-Lactamase in the HEK-CRE cell line (data not shown). In conclusion, the results are consistent with

HGPRBMY8 representing a functional GPCR analogous to known Gs coupled
receptors (Boss et al., 1996).

Demonstration of Cellular Expression:

5   HGPRBMY8 was tagged at the C-terminus using the Flag epitope and
inserted into the pcDNA3.1 hygro ™ expression vector, as described herein.
Immunocytochemistry of CHO-NFAT/CRE cell lines transfected with the Flag-
tagged HGPRBMY8 construct with FITC conjugated Anti Flag monoclonal antibody
demonstrated that HGPRBMY8 is indeed a cell surface receptor. The

10   immunocytochemistry also confirmed expression of the HGPRBMY8 in the CHO-
NFAT/CRE cell lines. Briefly, CHO-NFAT/CRE cell lines were transfected with
pcDNA3.1 hygro ™ / HGPRBMY8-Flag vector, fixed with 70% methanol, and
permeablized with 0.1%TritonX100. The cells were then blocked with 1% Serum and
incubated with a FITC conjugated Anti Flag monoclonal antibody at 1:50 dilution in

15   PBS-Triton. The cells were then washed several times with PBS-Triton, overlayed
with mounting solution, and fluorescent images were captured (see Figure 15A-D).
Figure 15 describes CHO-NFAT/CRE cell lines transfected with the pcDNA3.1
Hygro ™ / HGPRBMY8-FLAG mammalian expression vector subjected to
immunocytochemistry using an FITC conjugated Anti Flag monoclonal antibody.

20   Panel A shows the transfected CHO-NFAT/CRE cells under visual wavelengths, and
panel B shows the fluorescent emission of the same cells at 530 nm after illumination
with a mercury light source. The cell expression is clearly evident in panel B, and is
consistent with the HGPRBMY8 polypeptide representing a member of the GPCR
family. The control cell line, non-transfected CHO-NFAT /CREcell line, exhibited no

25   detectable background fluorescence (Figure 15). The HGPRBMY8 –FLAG tagged
expressing CHO-NFAT /CRE line exhibited specific plasma membrane expression as
indicated (Figure 15). These data provide clear evidence that HGPRBMY8 is
expressed in these cells and the majority of the protein is localized to the cell surface.
Cell surface localization in consistent with HGPRBM8 representing a 7

30   transmembrane domain containing GPCR. Taken together, the data indicate that
HGPRBMY8 is a cell surface GPCR that can function through increases in either
cAMP or Ca^{2+} signal transduction pathways via G alpha 15.

Screening Paradigm

        The Aurora Beta-Lactamase technology provides a clear path for identifying agonists and antagonists of the HGPRBMY8 polypeptide. Cell lines that

5   exhibit a range of constitutive coupling activity have been identified by sorting through HGPRBMY8 transfected cell lines using the FACS Vantage SE (see Figure 16). For example, cell lines have been sorted that have an intermediate level of orphan GPCR expression, which also correlates with an intermediate coupling response, using the LJL analyst. Such cell lines will provide the opportunity to

10   screen, indirectly, for both agonists and antogonists of HGPRBMY8 by looking for inhibitors that block the beta lactamase response, or agonists that increase the beta lactamase response. As described herein, modulating the expression level of beta lactamase directly correlates with the level of cleaved CCF2 substrate. For example, this screening paradigm has been shown to work for the identification of modulators

15   of a known GPCR, 5HT6, that couples through Adenylate Cyclase, in addition to, the identification of modulators of the 5HT2c GPCR, that couples through changes in [Ca^{2+}]i. The data shown below represent cell lines that have been engineered with the desired pattern of HGPRBMY8 expression to enable the identification of potent small molecule agonists and antagonists. HGPRBMY8 modulator screens may be carried

20   out using a variety of high throughput methods known in the art, though preferably using the fully automated Aurora UHTSS system. The untransfected CHO-NFAT/CRE cell line represents the relative background level of beta lactamase expression (Figure 16; panel a). Figure 16 describes several CHO-NFAT/CRE cell lines transfected with the pcDNA3.1 Hygro ^{TM} / HGPRBMY8 mammalian expression

25   vector isolated via FACS that had either intermediate or high beta lactamase expression levels of constitutive activation. Panel A shows untransfected CHO-NFAT/CRE cells prior to stimulation with 10 nM PMA, 1 μM Thapsigargin, and 10 μM Forskolin (- P/T/F). Panel B shows CHO-NFAT/CRE cells after stimulation with 10 nM PMA, 1 μM Thapsigargin, and 10 μM Forskolin (+ P/T/F). Panel C

30   shows a representative orphan GPCR (oGPCR) transfected CHO-NFAT/CRE cells that have an intermediate level of beta lactamase expression. Panel D shows a representative orphan GPCR transfected CHO-NFAT/CRE that have a high level of

beta lactamase expression. Following treatment with a cocktail of 10 nM PMA, 1μM
Thapsigargin, and 10μM Forskolin (Figure 16; P/T/F; panel b), the cells fully activate
the CRE-NFAT response element demonstrating the dynamic range of the assay.
Panel C (Figure 16) represents an orphan transfected CHO-NFAT/CRE cell line that
5   shows an intermediate level of beta lactamase expression post P/T/F stimulation,
while panel D (Figure 16) represents a orphan transfected CHO-NFAT/CRE cell line
that shows a high level of constitutive beta lactamase expression.

<div align="center">EXAMPLE 8</div>

<div align="center">G-PROTEIN COUPLED RECEPTOR PCR EXPRESSION PROFILING</div>

10   RNA quantification was performed using the Taqman real-time-PCR
fluorogenic assay. The Taqman assay is one of the most precise methods for assaying
the concentration of nucleic acid templates .

All cell lines were grown using standard conditions: RPMI 1640
supplemented with 10% fetal bovine serum, 100 IU/ml penicillin, 100 mg/ml
15   streptomycin, and 2 mM L-glutamine, 10 mM Hepes (all from GibcoBRL; Rockville,
MD). Eighty percent confluent cells were washed twice with phosphate-buffered
saline (GibcoBRL) and harvested using 0.25% trypsin (GibcoBRL). RNA was
prepared using the RNeasy Maxi Kit from Qiagen (Valencia, CA).

cDNA template for real-time PCR was generated using the Superscript
20   First Strand Synthesis system for RT-PCR.

SYBR Green real-time PCR reactions were prepared as follows: The
reaction mix consisted of 20 ng first strand cDNA; 50 nM Forward Primer; 50 nM
Reverse Primer; 0.75X SYBR Green I (Sigma); 1X SYBR Green PCR Buffer
(50mMTris-HCl pH8.3, 75mM KCl); 10%DMSO; 3mM MgCl$_2$; 300 M each dATP,
25   dGTP, dTTP, dCTP; 1 U Platinum Taq DNA Polymerase High Fidelity (Cat# 11304-
029; Life Technologies; Rockville, MD); 1:50 dilution; ROX (Life Technologies).
Real-time PCR was performed using an Applied Biosystems 5700 Sequence
Detection System. Conditions were 95C for 10 min (denaturation and activation of
Platinum Taq DNA Polymerase), 40 cycles of PCR (95C for 15 sec, 60C for 1 min).

PCR products are analyzed for uniform melting using an analysis algorithm built into the 5700 Sequence Detection System.

Forward primer: 745 GPCR84-2s: 5'-GCAGAGCACTCCTCCACTCT-3' (SEQ ID NO:34); and

5   Reverse primer: 746 GPCR84-2a: 5'-AGCAGGCAATCATGACAATC-3' (SEQ ID NO:35).

cDNA quantification used in the normalization of template quantity was performed using Taqman technology. Taqman reactions are prepared as follows: The reaction mix consisted of 20 ng first strand cDNA; 25 nM GAPDH-F3, Forward
10   Primer; 250 nM GAPDH-R1 Reverse Primer; 200 nM GAPDH-PVIC Taqman Probe (fluorescent dye labeled oligonucleotide primer); 1X Buffer A (Applied Biosystems); 5.5 mM MgCl2; 300 M dATP, dGTP, dTTP, dCTP; 1 U Amplitaq Gold (Applied Biosystems). GAPDH, D-glyceraldehyde -3-phosphate dehydrogenase, was used as control to normalize mRNA levels.

15   Real-time PCR was performed using an Applied Biosystems 7700 Sequence Detection System. Conditions were 95C for 10 min. (denaturation and activation of Amplitaq Gold), 40 cycles of PCR (95C for 15 sec, 60C for 1 min).

The sequences for the GAPDH oligonucleotides used in the Taqman reactions are as follows:

20   GAPDH-F3 -5'-AGCCGAGCCACATCGCT-3' (SEQ ID NO:60)

GAPDH-R1 -5'-GTGACCAGGCGCCCAATAC-3' (SEQ ID NO:61)

GAPDH-PVIC Taqman® Probe -VIC-5' -
CAAATCCGTTGACTCCGACCTTCACCTT-3' TAMRA (SEQ ID NO:62).

The Sequence Detection System generates a Ct (threshold cycle) value
25   that is used to calculate a concentration for each input cDNA template. cDNA levels for each gene of interest are normalized to GAPDH cDNA levels to compensate for variations in total cDNA quantity in the input sample. This is done by generating GAPDH Ct values for each cell line. Ct values for the gene of interest and GAPDH are inserted into a modified version of the Ct equation (Applied Biosystems Prism

7700 Sequence Detection System User Bulletin #2), which is used to calculate a GAPDH normalized relative cDNA level for each specific cDNA. The Ct equation is as follows: relative quantity of nucleic acid template $= 2^{Ct} = 2^{(Cta-Ctb)}$, where Cta=Ct target – Ct GAPDH, and Ctb = Ct reference – Ct GAPDH. (No reference cell line was used for the calculation of relative quantity; Ctb was defined as 21).

The Graph # of Table 1 corresponds to the tissue type position number of Figure 17. HGPRBMY8 (also known as GPCR84 or GPCR58) was found to have relatively low expression in the tumor cell lines assayed in the OCLP-1 (oncology cell line panel). HGPRBMY8 message appears to be especially scarce in breast tumor cell lines. The average HGPRBMY8 message message level in lung tumor cell lines is 2-3 fold lower than the average for other cell lines assayed.

TABLE 1

Graph #	Name	Tissue	Ct GAPDH	Ct GPCR84	dCt	ddCt	Quant.
1	AIN 4	breast	17.49	40	22.51	1.51	0.0E+00
2	AIN 4T	breast	17.15	36.8	19.65	-1.35	2.5E+00
3	AIN4/myc	breast	17.81	40	22.19	1.19	0.0E+00
4	BT-20	breast	17.9	36.15	18.25	-2.75	6.7E+00
5	BT-474	breast	17.65	38.34	20.69	-0.31	1.2E+00
6	BT-483	breast	17.45	35.6	18.15	-2.85	7.2E+00
7	BT-549	breast	17.55	38.21	20.66	-0.34	1.3E+00
8	DU4475	breast	18.1	40	21.9	0.9	0.0E+00
9	H3396	breast	18.04	36.71	18.67	-2.33	5.0E+00
10	HBL100	breast	17.02	37.16	20.14	-0.86	1.8E+00
11	Her2 MCF-7	breast	19.26	35.62	16.36	-4.64	2.5E+01
12	HS 578T	breast	17.83	37.28	19.45	-1.55	2.9E+00
13	MCF7	breast	17.83	40	22.17	1.17	0.0E+00
14	MCF-7/AdrR	breast	17.23	36.01	18.78	-2.22	4.7E+00

TABLE 1

Graph #	Name	Tissue	Ct GAPDH	Ct GPCR84	dCt	ddCt	Quant.
15	MDAH 2774	breast	16.87	35.24	18.37	-2.63	6.2E+00
16	MDA-MB-175-VII	breast	15.72	34.08	18.36	-2.64	6.2E+00
17	MDA-MB-231	breast	17.62	40	22.38	1.38	0.0E+00
18	MDA-MB-453	breast	17.9	37.57	19.67	-1.33	2.5E+00
19	MDA-MB-468	breast	17.49	37.58	20.09	-0.91	1.9E+00
20	Pat-21 R60	breast	35.59	40	4.41	-16.59	ND
21	SKBR3	breast	17.12	40	22.88	1.88	0.0E+00
22	T47D	breast	18.86	40	21.14	0.14	0.0E+00
23	UACC-812	breast	17.06	38.26	21.2	0.2	8.7E-01
24	ZR-75-1	breast	15.95	35.36	19.41	-1.59	3.0E+00
25	C-33A	cervical	17.49	36.96	19.47	-1.53	2.9E+00
26	Ca Ski	cervical	17.38	37.78	20.4	-0.6	1.5E+00
27	HeLa	cervical	17.59	40	22.41	1.41	0.0E+00
28	HT-3	cervical	17.42	35.69	18.27	-2.73	6.6E+00
29	ME-180	cervical	16.86	34.57	17.71	-3.29	9.8E+00
30	SiHa	cervical	18.07	40	21.93	0.93	0.0E+00
31	SW756	cervical	15.59	36.45	20.86	-0.14	1.1E+00
32	CACO-2	colon	17.56	40	22.44	1.44	0.0E+00
33	CCD-112Co	colon	18.03	40	21.97	0.97	0.0E+00
34	CCD-33Co	colon	17.07	39.44	22.37	1.37	3.9E-01
35	Colo 205	colon	18.02	40	21.98	0.98	0.0E+00
36	Colo 320DM	colon	17.01	40	22.99	1.99	0.0E+00
37	Colo201	colon	17.89	34.47	16.58	-4.42	2.1E+01
38	Cx-1	colon	18.79	40	21.21	0.21	0.0E+00

TABLE 1

Graph #	Name	Tissue	Ct GAPDH	Ct GPCR84	dCt	ddCt	Quant.
39	ddH2O	colon	40	40	0	-21	ND
40	HCT116	colon	17.59	36.22	18.63	-2.37	5.2E+00
41	HCT116/epo5	colon	17.71	36.42	18.71	-2.29	4.9E+00
42	HCT116/ras	colon	17.18	40	22.82	1.82	0.0E+00
43	HCT116/TX15 CR	colon	17.36	36.91	19.55	-1.45	2.7E+00
44	HCT116/vivo	colon	17.7	37.01	19.31	-1.69	3.2E+00
45	HCT116/VM46	colon	17.87	37.55	19.68	-1.32	2.5E+00
46	HCT116/VP35	colon	17.3	40	22.7	1.7	0.0E+00
47	HCT-8	colon	17.44	36.86	19.42	-1.58	3.0E+00
48	HT-29	colon	17.9	40	22.1	1.1	0.0E+00
49	LoVo	colon	17.64	40	22.36	1.36	0.0E+00
50	LS 174T	colon	17.93	40	22.07	1.07	0.0E+00
51	LS123	colon	17.65	36.05	18.4	-2.6	6.1E+00
52	MIP	colon	16.92	35.65	18.73	-2.27	4.8E+00
53	SK-CO-1	colon	17.75	39.84	22.09	1.09	4.7E-01
54	SW1417	colon	17.22	39.11	21.89	0.89	5.4E-01
55	SW403	colon	18.39	40	21.61	0.61	0.0E+00
56	SW480	colon	17	40	23	2	0.0E+00
57	SW620	colon	17.16	40	22.84	1.84	0.0E+00
58	SW837	colon	18.35	37.65	19.3	-1.7	3.2E+00
59	T84	colon	16.44	40	23.56	2.56	0.0E+00
60	CCD-18Co	colon, fibroblast	17.19	40	22.81	1.81	0.0E+00
61	HT-1080	fibrosarcoma	17.16	40	22.84	1.84	0.0E+00

TABLE 1

Graph #	Name	Tissue	Ct GAPDH	Ct GPCR84	dCt	ddCt	Quant.
62	CCRF-CEM	leukemia	17.07	40	22.93	1.93	0.0E+00
63	HL-60	leukemia	17.54	40	22.46	1.46	0.0E+00
64	K562	leukemia	18.42	40	21.58	0.58	0.0E+00
65	A-427	lung	18	40	22	1	0.0E+00
66	A549	lung	17.63	37.06	19.43	-1.57	3.0E+00
67	Calu-3	lung	18.09	37.38	19.29	-1.71	3.3E+00
68	Calu-6	lung	16.62	40	23.38	2.38	0.0E+00
69	ChaGo-K-1	lung	17.79	37.16	19.37	-1.63	3.1E+00
70	DMS 114	lung	18.14	40	21.86	0.86	0.0E+00
71	LX-1	lung	18.17	40	21.83	0.83	0.0E+00
72	MRC-5	lung	17.3	40	22.7	1.7	0.0E+00
73	MSTO-211H	lung	16.81	40	23.19	2.19	0.0E+00
74	NCI-H596	lung	17.73	40	22.27	1.27	0.0E+00
75	SHP-77	lung	18.66	40	21.34	0.34	0.0E+00
76	Sk-LU-1	lung	15.81	35.83	20.02	-0.98	2.0E+00
77	SK-MES-1	lung	17.1	36.33	19.23	-1.77	3.4E+00
78	SW1271	lung	16.45	40	23.55	2.55	0.0E+00
79	SW1573	lung	17.14	40	22.86	1.86	0.0E+00
80	SW900	lung	18.17	40	21.83	0.83	0.0E+00
81	Hs 294T	melanoma	17.73	35.38	17.65	-3.35	1.0E+01
82	A2780/DDP-R	ovarian	21.51	40	18.49	-2.51	0.0E+00
83	A2780/DDP-S	ovarian	17.89	35.73	17.84	-3.16	8.9E+00
84	A2780/epo5	ovarian	17.54	35.12	17.58	-3.42	1.1E+01
85	A2780/TAX-R	ovarian	18.4	38.33	19.93	-1.07	2.1E+00
86	A2780/TAX-S	ovarian	17.83	40	22.17	1.17	0.0E+00

TABLE 1

Graph #	Name	Tissue	Ct GAPDH	Ct GPCR84	dCt	ddCt	Quant.
87	Caov-3	ovarian	15.5	35.35	19.85	-1.15	2.2E+00
88	ES-2	ovarian	17.22	40	22.78	1.78	0.0E+00
89	HOC-76	ovarian	34.3	40	5.7	-15.3	ND
90	OVCAR-3	ovarian	17.09	36.66	19.57	-1.43	2.7E+00
91	PA-1	ovarian	17.33	40	22.67	1.67	0.0E+00
92	SW 626	ovarian	16.94	40	23.06	2.06	0.0E+00
93	UPN251	ovarian	17.69	36.75	19.06	-1.94	3.8E+00
94	LNCAP	prostate	18.17	40	21.83	0.83	0.0E+00
95	PC-3	prostate	17.25	40	22.75	1.75	0.0E+00
96	A431	squamous	19.85	40	20.15	-0.85	0.0E+00

EXAMPLE 9

PHAGE DISPLAY METHODS FOR IDENTIFYING PEPTIDE LIGANDS OR
MODULATORS OF ORPHAN GPCRS

Library Construction

Two HGPRBMY libraries were used for identifying peptides that may function as modulators. Specifically, a 15-mer library was used to identify peptides that may function as agonists or antagonists. The 15-mer library is an aliquot of the 15-mer library originally constructed by G.P. Smith (Scott, JK and Smith, GP. 1990, Science 249:386-390). A 40-mer library was used for identifying natural ligands and constructed essentially as previously described, using an M13 phage library displaying random 38-amino acid peptides as a source of novel sequences with affinity to selected targets (BK Kay, et al. 1993, Gene 128:59-65). This method for constructing the 40-mer library was followed with the exception that a 15 base pair complementary region was used to anneal the two oligonucleotides, as opposed to 6, 9, or 12 base pairs, as described below.

The oligos used are: Oligo 1: 5'-
CGAAGCGTAAGGGCCCAGCCGGCCNN (BNNx19) BCCGGGTCCGGGCGGC -
3' (SEQ ID NO:63) and Oligo2: 5'- AAAAGGAAAAAAGCGGCCGC (VNNx20)
GCCGCCCGGACCCGG-3' (SEQ ID NO:64), where N= A+G+C+T and B = C+G+T

5   and V=C+A+G.

The oligos were are annealed via their 15 base pair complimentary
sequences which encode a constant ProGlyProGlyGly (SEQ ID NO:65) pentapeptide
sequence between the random 20 amino acid segments, and then extended by standard
procedure using Klenow enzyme. This was followed by endonuclease digestion using

10   Sfi1 and Not1 enzymes and ligation to Sfi1 and Not1 cleaved pCantab5E
(Pharmacia). The ligation mixture was electroporated into *E. coli* XL1Blue and phage
clones were essentially generated as suggested by the manufacturer (Pharmacia) for
making ScFv antibody libraries in pCantab5E.

Sequencing Bound Phage

15   Standard procedures commonly known in the art were used. Phage in
eluates were infected into *E. coli* host strain (TG1 for the 15-mer library; XL1Blue for
the 40-mer library) and plated for single colonies. Colonies were grown in liquid and
sequenced by standard procedure which involved: 1) generating PCR product with
suitable primers of the library segments in the phage genome (15-mer library) or

20   pCantab5E (40-mer library); and 2) sequencing PCR products using one primer of
each PCR primer pair. Sequences were visually inspected or were inspected by using
the Vector NTI alignment tool.

Peptide Modulators Of The Present Invention

The following serve as non-limiting examples of HGPRBMY8 peptide

25   modulators:

GDFWYEACESSCAFW	(SEQ ID NO:66)
LEWGSDVFYDVYDCC	(SEQ ID NO:67)
CLRSGTGCAFQLYRF	(SEQ ID NO:68)
NNFPCLRSGRNCDAG	(SEQ ID NO:69)
RIVPNGYFNVHGRSL	(SEQ ID NO:70)
RIDSCAKYFLRSCD	(SEQ ID NO:71)

Peptide Synthesis

Peptides were synthesized on Fmoc-Knorr amide resin [N-(9-
fluorenyl)methoxycarbonyl-Knorr amide-resin, Midwest Biotech, Fishers, IN] with an
Applied Biosystems (Foster City, CA) model 433A synthesizer and the *FastMoc*

5   chemistry protocol (0.25mmol scale) supplied with the instrument. Amino acids were
double coupled as their N-alpha-Fmoc- derivatives and reactive side chains were
protected as follows: Asp, Glu: t-Butyl ester (OtBu); Ser, Thr, Tyr: t-Butyl ether
(tBu); Asn, Cys, Gln, His: Triphenylmethyl (Trt); Lys, Trp: t-Butyloxycarbonyl
(Boc); Arg: 2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl (Pbf). After the final

10   double coupling cycle, the N-terminal Fmoc group was removed by the multi-step
treatment with piperidine in N-Methylpyrrolidone as described by the manufacturer.
The N-terminal free amines were then treated with 10% acetic anhydride, 5%
Diisopropylamine in N-Methylpyrrolidone to yield the N-acetyl-derivative. The
protected peptidyl-resins were simultaneously deprotected and removed from the

15   resin by standard methods. The lyophilized peptides were purified on C_{18} to apparent
homogeneity as judged by RP-HPLC analysis. Predicted peptide molecular weights
were verified by electrospray mass spectrometry (J. Biol. Chem. vol. 273, pp.12041-
12046, 1998).

Cyclic analogs were prepared from the crude linear products. The

20   cystine disulfide was formed using one of the following methods:
Method 1:

A sample of the crude peptide was dissolved in water at a
concentration of 0.5 mg/mL and the pH adjusted to 8.5 with NH_4OH. The reaction
was stirred at room temperature, and monitored by RP-HPLC. Once complete, the

25   reaction was brought to pH 4 with acetic acid and lyophilized. The product was
purified and characterized as above.
Method 2

A sample of the crude peptide was dissolved at a concentration of
0.5mg/mL in 5% acetic acid. The pH was adjusted to 6.0 with NH_4OH. DMSO (20%

30   by volume) was added and the reaction was stirred overnight. After analytical RP-
HPLC analysis, the reaction was diluted with water and triple lyophilized to remove

DMSO. The crude product was purified by preparative RP-HPLC. (JACS. vol. 113, 6657, 1991).

Assessing Affect of Peptides on GPCR Function

The effect of any one of these peptides on the function of the GPCR of

5   the present invention may be determined by adding an effective amount of each

peptide to each functional assay. Representative functional assays are described more

specifically herein, particularly Example 7.

Uses Of The Peptide Modulators Of The Present Invention

The aforementioned peptides of the present invention are useful for a

10   variety of purposes, though most notably for modulating the function of the GPCR of

the present invention, and potentially with other GPCRs of the same G-protein

coupled receptor subclass (e.g., peptide receptors, adrenergic receptors, purinergic

receptors, etc.), and/or other subclasses known in the art. For example, the peptide

modulators of the present invention may be useful as HGPRBMY8 agonists.

15   Alternatively, the peptide modulators of the present invention may be useful as

HGPRBMY8 antagonists of the present invention. In addition, the peptide

modulators of the present invention may be useful as competitive inhibitors of the

HGPRBMY8 cognate ligand(s), or may be useful as non-competitive inhibitors of the

HGPRBMY8 cognate ligand(s).

20   Furthermore, the peptide modulators of the present invention may be

useful in assays designed to either deorphan the HGPRBMY8 polypeptide of the

present invention, or to identify other agonists or antagonists of the HGPRBMY8

polypeptide of the present invention, particularly small molecule modulators.

25   EXAMPLE 10

METHOD OF CREATING N- AND C- TERMINAL DELETION MUTANTS

CORRESPONDING TO THE HGPRBMY8 POLYPEPTIDE

As described elsewhere herein, the present invention encompasses the

creation of N- and C-terminal deletion mutants, in addition to any combination of N-

30   and C-terminal deletions thereof, corresponding to the HGPRBMY8 polypeptide of

the present invention. A number of methods are available to one skilled in the art for

creating such mutants. Such methods may include a combination of PCR

amplification and gene cloning methodology. Although one of skill in the art of molecular biology, through the use of the teachings provided or referenced herein, and/or otherwise known in the art as standard methods, could readily create each deletion mutants of the present invention, exemplary methods are described below.

Briefly, using the isolated cDNA clone encoding the full-length HGPRBMY8 polypeptide sequence, appropriate primers of about 15-25 nucleotides derived from the desired 5' and 3' positions of SEQ ID NO:1 may be designed to PCR amplify, and subsequently clone, the intended N- and/or C-terminal deletion mutant. Such primers could comprise, for example, an initiation and stop codon for the 5' and 3' primer, respectively. Such primers may also comprise restriction sites to facilitate cloning of the deletion mutant post amplification. Moreover, the primers may comprise additional sequences, such as, for example, flag-tag sequences, kozac sequences, or other sequences discussed and/or referenced herein.

For example, in the case of the T36 to P508 N-terminal deletion mutant, the following primers could be used to amplify a cDNA fragment corresponding to this deletion mutant:

5' Primer   5'-GCAGCA <u>GCGGCCGC</u> ACCGTGCTGGTTATCTTCCTCGCCG –3' (SEQ ID NO:72)
                          NotI

3' Primer   5'- GCAGCA <u>GTCGAC</u> AGGAAAAGTAGCAGAATCGTAGG -3' (SEQ ID NO:73)
                          SalI

For example, in the case of the M1 to Y454 C-terminal deletion mutant, the following primers could be used to amplify a cDNA fragment corresponding to this deletion mutant:

5' Primer   5'- GCAGCA <u>GCGGCCGC</u> ATGACGTCCACCTGCACCAACAGC –3' (SEQ ID NO:74)
                          NotI

3' Primer   5'- GCAGCA <u>GTCGAC</u> ATAGACATAGGGGTGGATGCAGCAC -3' (SEQ ID NO:75)
                          SalI

[0102]   Representative PCR amplification conditions are provided below,

although the skilled artisan would appreciate that other conditions may be required for efficient amplification. A 100 μl PCR reaction mixture may be prepared using 10ng of the template DNA (cDNA clone of HGPRBMY8), 200 μM 4dNTPs, 1μM primers, 0.25U Taq DNA polymerase (PE), and standard Taq DNA polymerase buffer.

5   Typical PCR cycling condition are as follows:

> 20-25 cycles: 45 sec, 93 degrees
>
> 2 min, 50 degrees
>
> 2 min, 72 degrees
>
> 1 cycle: 10 min, 72 degrees

10   After the final extension step of PCR, 5U Klenow Fragment may be added and incubated for 15 min at 30 degrees.

Upon digestion of the fragment with the NotI and SalI restriction enzymes, the fragment could be cloned into an appropriate expression and/or cloning vector which has been similarly digested (e.g., pSport1, among others). The skilled

15   artisan would appreciate that other plasmids could be equally substituted, and may be desirable in certain circumstances. The digested fragment and vector are then ligated using a DNA ligase, and then used to transform competent *E.coli* cells using methods provided herein and/or otherwise known in the art.

The 5' primer sequence for amplifying any additional N-terminal

20   deletion mutants may be determined by reference to the following formula:

$$(S+(X * 3)) \text{ to } ((S+(X * 3))+25),$$

wherein 'S' is equal to the nucleotide position of the initiating start codon of the HGPRBMY8 gene (SEQ ID NO:1), and 'X' is equal to the most N-terminal amino acid of the intended N-terminal deletion mutant. The first term provides the start 5'

25   nucleotide position of the 5' primer, while the second term provides the end 3' nucleotide position of the 5' primer corresponding to sense strand of SEQ ID NO:1. Once the corresponding nucleotide positions of the primer are determined, the final nucleotide sequence may be created by the addition of applicable restriction site sequences to the 5' end of the sequence, for example. As referenced herein, the

30   addition of other sequences to the 5' primer may be desired in certain circumstances (e.g., kozac sequences, etc.).

The 3' primer sequence for amplifying any additional N-terminal deletion mutants may be determined by reference to the following formula:

(S+(X * 3)) to ((S+(X * 3))-25),

wherein 'S' is equal to the nucleotide position of the initiating start codon of the

5   HGPRBMY8 gene (SEQ ID NO:1), and 'X' is equal to the most C-terminal amino acid of the intended N-terminal deletion mutant. The first term provides the start 5' nucleotide position of the 3' primer, while the second term provides the end 3' nucleotide position of the 3' primer corresponding to the anti-sense strand of SEQ ID NO:1. Once the corresponding nucleotide positions of the primer are determined, the

10   final nucleotide sequence may be created by the addition of applicable restriction site sequences to the 5' end of the sequence, for example. As referenced herein, the addition of other sequences to the 3' primer may be desired in certain circumstances (e.g., stop codon sequences, etc.). The skilled artisan would appreciate that modifications of the above nucleotide positions may be necessary for optimizing PCR

15   amplification.

The same general formulas provided above may be used in identifying the 5' and 3' primer sequences for amplifying any C-terminal deletion mutant of the present invention. Moreover, the same general formulas provided above may be used in identifying the 5' and 3' primer sequences for amplifying any combination of N-

20   terminal and C-terminal deletion mutant of the present invention. The skilled artisan would appreciate that modifications of the above nucleotide positions may be necessary for optimizing PCR amplification.

In preferred embodiments, the following N-terminal HGPRBMY8 deletion polypeptides are encompassed by the present invention: M1-P508, T2-P508,

25   S3-P508, T4-P508, C5-P508, T6-P508, N7-P508, S8-P508, T9-P508, R10-P508, E11-P508, S12-P508, N13-P508, S14-P508, S15-P508, H16-P508, T17-P508, C18-P508, M19-P508, P20-P508, L21-P508, S22-P508, K23-P508, M24-P508, P25-P508, I26-P508, S27-P508, L28-P508, A29-P508, H30-P508, G31-P508, I32-P508, I33-P508, R34-P508, S35-P508, T36-P508, V37-P508, L38-P508, V39-P508, I40-P508, F41-

30   P508, L42-P508, A43-P508, A44-P508, S45-P508, F46-P508, V47-P508, G48-P508, N49-P508, I50-P508, V51-P508, L52-P508, A53-P508, L54-P508, V55-P508, L56-

102

P508, Q57-P508, R58-P508, K59-P508, P60-P508, Q61-P508, L62-P508, L63-P508, Q64-P508, V65-P508, T66-P508, N67-P508, R68-P508, F69-P508, I70-P508, F71-P508, N72-P508, L73-P508, L74-P508, V75-P508, T76-P508, D77-P508, L78-P508, L79-P508, Q80-P508, I81-P508, S82-P508, L83-P508, V84-P508, A85-P508, P86-

5   P508, W87-P508, V88-P508, V89-P508, A90-P508, T91-P508, S92-P508, V93-P508, P94-P508, L95-P508, F96-P508, W97-P508, P98-P508, L99-P508, N100-P508, S101-P508, H102-P508, F103-P508, C104-P508, T105-P508, A106-P508, L107-P508, V108-P508, S109-P508, L110-P508, T111-P508, H112-P508, L113-P508, F114-P508, A115-P508, F116-P508, A117-P508, S118-P508, V119-P508, N120-

10   P508, T121-P508, I122-P508, V123-P508, L124-P508, V125-P508, S126-P508, V127-P508, D128-P508, R129-P508, Y130-P508, L131-P508, S132-P508, I133-P508, I134-P508, H135-P508, P136-P508, L137-P508, S138-P508, Y139-P508, P140-P508, S141-P508, K142-P508, M143-P508, T144-P508, Q145-P508, R146-P508, R147-P508, G148-P508, Y149-P508, L150-P508, L151-P508, L152-P508,

15   Y153-P508, G154-P508, T155-P508, W156-P508, I157-P508, V158-P508, A159-P508, I160-P508, L161-P508, Q162-P508, S163-P508, T164-P508, P165-P508, P166-P508, L167-P508, Y168-P508, G169-P508, W170-P508, G171-P508, Q172-P508, A173-P508, A174-P508, F175-P508, D176-P508, E177-P508, R178-P508, N179-P508, A180-P508, L181-P508, C182-P508, S183-P508, M184-P508, I185-

20   P508, W186-P508, G187-P508, A188-P508, S189-P508, P190-P508, S191-P508, Y192-P508, T193-P508, I194-P508, L195-P508, S196-P508, V197-P508, V198-P508, S199-P508, F200-P508, I201-P508, V202-P508, I203-P508, P204-P508, L205-P508, I206-P508, V207-P508, M208-P508, I209-P508, A210-P508, C211-P508, Y212-P508, S213-P508, V214-P508, V215-P508, F216-P508, C217-P508, A218-

25   P508, A219-P508, R220-P508, R221-P508, Q222-P508, H223-P508, A224-P508, L225-P508, L226-P508, Y227-P508, N228-P508, V229-P508, K230-P508, R231-P508, H232-P508, S233-P508, L234-P508, E235-P508, V236-P508, R237-P508, V238-P508, K239-P508, D240-P508, C241-P508, V242-P508, E243-P508, N244-P508, E245-P508, D246-P508, E247-P508, E248-P508, G249-P508, A250-P508,

30   E251-P508, K252-P508, K253-P508, E254-P508, E255-P508, F256-P508, Q257-P508, D258-P508, E259-P508, S260-P508, E261-P508, F262-P508, R263-P508, R264-P508, Q265-P508, H266-P508, E267-P508, G268-P508, E269-P508, V270-

P508, K271-P508, A272-P508, K273-P508, E274-P508, G275-P508, R276-P508,

M277-P508, E278-P508, A279-P508, K280-P508, D281-P508, G282-P508, S283-

P508, L284-P508, K285-P508, A286-P508, K287-P508, E288-P508, G289-P508,

S290-P508, T291-P508, G292-P508, T293-P508, S294-P508, E295-P508, S296-

5   P508, S297-P508, V298-P508, E299-P508, A300-P508, R301-P508, G302-P508,

S303-P508, E304-P508, E305-P508, V306-P508, R307-P508, E308-P508, S309-

P508, S310-P508, T311-P508, V312-P508, A313-P508, S314-P508, D315-P508,

G316-P508, S317-P508, M318-P508, E319-P508, G320-P508, K321-P508, E322-

P508, G323-P508, S324-P508, T325-P508, K326-P508, V327-P508, E328-P508,

10   E329-P508, N330-P508, S331-P508, M332-P508, K333-P508, A334-P508, D335-

P508, K336-P508, G337-P508, R338-P508, T339-P508, E340-P508, V341-P508,

N342-P508, Q343-P508, C344-P508, S345-P508, I346-P508, D347-P508, L348-

P508, G349-P508, E350-P508, D351-P508, D352-P508, M353-P508, E354-P508,

F355-P508, G356-P508, E357-P508, D358-P508, D359-P508, I360-P508, N361-

15   P508, F362-P508, S363-P508, E364-P508, D365-P508, D366-P508, V367-P508,

E368-P508, A369-P508, V370-P508, N371-P508, I372-P508, P373-P508, E374-

P508, S375-P508, L376-P508, P377-P508, P378-P508, S379-P508, R380-P508,

R381-P508, N382-P508, S383-P508, N384-P508, S385-P508, N386-P508, P387-

P508, P388-P508, L389-P508, P390-P508, R391-P508, C392-P508, Y393-P508,

20   Q394-P508, C395-P508, K396-P508, A397-P508, A398-P508, K399-P508, V400-

P508, I401-P508, F402-P508, I403-P508, I404-P508, I405-P508, F406-P508, S407-

P508, Y408-P508, V409-P508, L410-P508, S411-P508, L412-P508, G413-P508,

P414-P508, Y415-P508, C416-P508, F417-P508, L418-P508, A419-P508, V420-

P508, L421-P508, A422-P508, V423-P508, W424-P508, V425-P508, D426-P508,

25   V427-P508, E428-P508, T429-P508, Q430-P508, V431-P508, P432-P508, Q433-

P508, W434-P508, V435-P508, I436-P508, T437-P508, I438-P508, I439-P508, I440-

P508, W441-P508, L442-P508, F443-P508, F444-P508, L445-P508, Q446-P508,

C447-P508, C448-P508, I449-P508, H450-P508, P451-P508, Y452-P508, V453-

P508, Y454-P508, G455-P508, Y456-P508, M457-P508, H458-P508, K459-P508,

30   T460-P508, I461-P508, K462-P508, K463-P508, E464-P508, I465-P508, Q466-P508,

D467-P508, M468-P508, L469-P508, K470-P508, K471-P508, F472-P508, F473-

P508, C474-P508, K475-P508, E476-P508, K477-P508, P478-P508, P479-P508,

104

K480-P508, E481-P508, D482-P508, S483-P508, H484-P508, P485-P508, D486-
P508, L487-P508, P488-P508, G489-P508, T490-P508, E491-P508, G492-P508,
G493-P508, T494-P508, E495-P508, G496-P508, K497-P508, I498-P508, V499-
P508, P500-P508, S501-P508, and/or Y502-P508 of SEQ ID NO:2. Polynucleotide

5   sequences encoding these polypeptides are also provided. The present invention also
encompasses the use of these N-terminal HGPRBMY8 deletion polypeptides as
immunogenic and/or antigenic epitopes as described elsewhere herein.

In preferred embodiments, the following C-terminal HGPRBMY8
deletion polypeptides are encompassed by the present invention: M1-P508, M1-F507,

10   M1-T506, M1-A505, M1-S504, M1-D503, M1-Y502, M1-S501, M1-P500, M1-
V499, M1-I498, M1-K497, M1-G496, M1-E495, M1-T494, M1-G493, M1-G492,
M1-E491, M1-T490, M1-G489, M1-P488, M1-L487, M1-D486, M1-P485, M1-
H484, M1-S483, M1-D482, M1-E481, M1-K480, M1-P479, M1-P478, M1-K477,
M1-E476, M1-K475, M1-C474, M1-F473, M1-F472, M1-K471, M1-K470, M1-

15   L469, M1-M468, M1-D467, M1-Q466, M1-I465, M1-E464, M1-K463, M1-K462,
M1-I461, M1-T460, M1-K459, M1-H458, M1-M457, M1-Y456, M1-G455, M1-
Y454, M1-V453, M1-Y452, M1-P451, M1-H450, M1-I449, M1-C448, M1-C447,
M1-Q446, M1-L445, M1-F444, M1-F443, M1-L442, M1-W441, M1-I440, M1-I439,
M1-I438, M1-T437, M1-I436, M1-V435, M1-W434, M1-Q433, M1-P432, M1-V431,

20   M1-Q430, M1-T429, M1-E428, M1-V427, M1-D426, M1-V425, M1-W424, M1-
V423, M1-A422, M1-L421, M1-V420, M1-A419, M1-L418, M1-F417, M1-C416,
M1-Y415, M1-P414, M1-G413, M1-L412, M1-S411, M1-L410, M1-V409, M1-
Y408, M1-S407, M1-F406, M1-I405, M1-I404, M1-I403, M1-F402, M1-I401, M1-
V400, M1-K399, M1-A398, M1-A397, M1-K396, M1-C395, M1-Q394, M1-Y393,

25   M1-C392, M1-R391, M1-P390, M1-L389, M1-P388, M1-P387, M1-N386, M1-S385,
M1-N384, M1-S383, M1-N382, M1-R381, M1-R380, M1-S379, M1-P378, M1-P377,
M1-L376, M1-S375, M1-E374, M1-P373, M1-I372, M1-N371, M1-V370, M1-A369,
M1-E368, M1-V367, M1-D366, M1-D365, M1-E364, M1-S363, M1-F362, M1-
N361, M1-I360, M1-D359, M1-D358, M1-E357, M1-G356, M1-F355, M1-E354,

30   M1-M353, M1-D352, M1-D351, M1-E350, M1-G349, M1-L348, M1-D347, M1-
I346, M1-S345, M1-C344, M1-Q343, M1-N342, M1-V341, M1-E340, M1-T339,
M1-R338, M1-G337, M1-K336, M1-D335, M1-A334, M1-K333, M1-M332, M1-

S331, M1-N330, M1-E329, M1-E328, M1-V327, M1-K326, M1-T325, M1-S324, M1-G323, M1-E322, M1-K321, M1-G320, M1-E319, M1-M318, M1-S317, M1-G316, M1-D315, M1-S314, M1-A313, M1-V312, M1-T311, M1-S310, M1-S309, M1-E308, M1-R307, M1-V306, M1-E305, M1-E304, M1-S303, M1-G302, M1-

5   R301, M1-A300, M1-E299, M1-V298, M1-S297, M1-S296, M1-E295, M1-S294, M1-T293, M1-G292, M1-T291, M1-S290, M1-G289, M1-E288, M1-K287, M1-A286, M1-K285, M1-L284, M1-S283, M1-G282, M1-D281, M1-K280, M1-A279, M1-E278, M1-M277, M1-R276, M1-G275, M1-E274, M1-K273, M1-A272, M1-K271, M1-V270, M1-E269, M1-G268, M1-E267, M1-H266, M1-Q265, M1-R264,

10   M1-R263, M1-F262, M1-E261, M1-S260, M1-E259, M1-D258, M1-Q257, M1-F256, M1-E255, M1-E254, M1-K253, M1-K252, M1-E251, M1-A250, M1-G249, M1-E248, M1-E247, M1-D246, M1-E245, M1-N244, M1-E243, M1-V242, M1-C241, M1-D240, M1-K239, M1-V238, M1-R237, M1-V236, M1-E235, M1-L234, M1-S233, M1-H232, M1-R231, M1-K230, M1-V229, M1-N228, M1-Y227, M1-L226,

15   M1-L225, M1-A224, M1-H223, M1-Q222, M1-R221, M1-R220, M1-A219, M1-A218, M1-C217, M1-F216, M1-V215, M1-V214, M1-S213, M1-Y212, M1-C211, M1-A210, M1-I209, M1-M208, M1-V207, M1-I206, M1-L205, M1-P204, M1-I203, M1-V202, M1-I201, M1-F200, M1-S199, M1-V198, M1-V197, M1-S196, M1-L195, M1-I194, M1-T193, M1-Y192, M1-S191, M1-P190, M1-S189, M1-A188, M1-G187,

20   M1-W186, M1-I185, M1-M184, M1-S183, M1-C182, M1-L181, M1-A180, M1-N179, M1-R178, M1-E177, M1-D176, M1-F175, M1-A174, M1-A173, M1-Q172, M1-G171, M1-W170, M1-G169, M1-Y168, M1-L167, M1-P166, M1-P165, M1-T164, M1-S163, M1-Q162, M1-L161, M1-I160, M1-A159, M1-V158, M1-I157, M1-W156, M1-T155, M1-G154, M1-Y153, M1-L152, M1-L151, M1-L150, M1-Y149,

25   M1-G148, M1-R147, M1-R146, M1-Q145, M1-T144, M1-M143, M1-K142, M1-S141, M1-P140, M1-Y139, M1-S138, M1-L137, M1-P136, M1-H135, M1-I134, M1-I133, M1-S132, M1-L131, M1-Y130, M1-R129, M1-D128, M1-V127, M1-S126, M1-V125, M1-L124, M1-V123, M1-I122, M1-T121, M1-N120, M1-V119, M1-S118, M1-A117, M1-F116, M1-A115, M1-F114, M1-L113, M1-H112, M1-T111, M1-

30   L110, M1-S109, M1-V108, M1-L107, M1-A106, M1-T105, M1-C104, M1-F103, M1-H102, M1-S101, M1-N100, M1-L99, M1-P98, M1-W97, M1-F96, M1-L95, M1-P94, M1-V93, M1-S92, M1-T91, M1-A90, M1-V89, M1-V88, M1-W87, M1-P86,

M1-A85, M1-V84, M1-L83, M1-S82, M1-I81, M1-Q80, M1-L79, M1-L78, M1-D77,
M1-T76, M1-V75, M1-L74, M1-L73, M1-N72, M1-F71, M1-I70, M1-F69, M1-R68,
M1-N67, M1-T66, M1-V65, M1-Q64, M1-L63, M1-L62, M1-Q61, M1-P60, M1-
K59, M1-R58, M1-Q57, M1-L56, M1-V55, M1-L54, M1-A53, M1-L52, M1-V51,

5   M1-I50, M1-N49, M1-G48, M1-V47, M1-F46, M1-S45, M1-A44, M1-A43, M1-L42,
M1-F41, M1-I40, M1-V39, M1-L38, M1-V37, M1-T36, M1-S35, M1-R34, M1-I33,
M1-I32, M1-G31, M1-H30, M1-A29, M1-L28, M1-S27, M1-I26, M1-P25, M1-M24,
M1-K23, M1-S22, M1-L21, M1-P20, M1-M19, M1-C18, M1-T17, M1-H16, M1-
S15, M1-S14, M1-N13, M1-S12, M1-E11, M1-R10, M1-T9, M1-S8, and/or M1-N7

10   of SEQ ID NO:2. Polynucleotide sequences encoding these polypeptides are also
provided. The present invention also encompasses the use of these C-terminal
HGPRBMY8 deletion polypeptides as immunogenic and/or antigenic epitopes as
described elsewhere herein.

          Alternatively, preferred polypeptides of the present invention may

15   comprise polypeptide sequences corresponding to, for example, internal regions of the
HGPRBMY8 polypeptide (e.g., any combination of both N- and C- terminal
HGPRBMY8 polypeptide deletions) of SEQ ID NO:2. For example, internal regions
could be defined by the equation: amino acid NX to amino acid CX, wherein NX
refers to any N-terminal deletion polypeptide amino acid of HGPRBMY8 (SEQ ID

20   NO:2), and where CX refers to any C-terminal deletion polypeptide amino acid of
HGPRBMY8 (SEQ ID NO:2). Polynucleotides encoding these polypeptides are also
provided. The present invention also encompasses the use of these polypeptides as an
immunogenic and/or antigenic epitope as described elsewhere herein.

<div align="center">EXAMPLE 11</div>

25   METHOD OF ENHANCING THE BIOLOGICAL ACTIVITY/FUNCTIONAL
CHARACTERISTICS OF INVENTION THROUGH MOLECULAR EVOLUTION.

          Although many of the most biologically active proteins known are
highly effective for their specified function in an organism, they often possess
characteristics that make them undesirable for transgenic, therapeutic, pharmaceutical,

30   and/or industrial applications. Among these traits, a short physiological half-life is
the most prominent problem, and is present either at the level of the protein, or the

<div align="center">107</div>

level of the proteins mRNA. The ability to extend the half-life, for example, would be particularly important for a proteins use in gene therapy, transgenic animal production, the bioprocess production and purification of the protein, and use of the protein as a chemical modulator among others. Therefore, there is a need to identify

5   novel variants of isolated proteins possessing characteristics which enhance their application as a therapeutic for treating diseases of animal origin, in addition to the proteins applicability to common industrial and pharmaceutical applications.

Thus, one aspect of the present invention relates to the ability to enhance specific characteristics of invention through directed molecular evolution.

10   Such an enhancement may, in a non-limiting example, benefit the inventions utility as an essential component in a kit, the inventions physical attributes such as its solubility, structure, or codon optimization, the inventions specific biological activity, including any associated enzymatic activity, the proteins enzyme kinetics, the proteins Ki, Kcat, Km, Vmax, Kd, protein-protein activity, protein-DNA binding activity,

15   antagonist/inhibitory activity (including direct or indirect interaction), agonist activity (including direct or indirect interaction), the proteins antigenicity (e.g., where it would be desirable to either increase or decrease the antigenic potential of the protein), the immunogenicity of the protein, the ability of the protein to form dimers, trimers, or multimers with either itself or other proteins, the antigenic efficacy of the invention,

20   including its subsequent use a preventative treatment for disease or disease states, or as an effector for targeting diseased genes. Moreover, the ability to enhance specific characteristics of a protein may also be applicable to changing the characterized activity of an enzyme to an activity completely unrelated to its initially characterized activity. Other desirable enhancements of the invention would be specific to each

25   individual protein, and would thus be well known in the art and contemplated by the present invention.

For example, an engineered G-protein coupled receptor may be constitutively active upon binding of its cognate ligand. Alternatively, an engineered G-protein coupled receptor may be constitutively active in the absence of ligand

30   binding. In yet another example, an engineered GPCR may be capable of being activated with less than all of the regulatory factors and/or conditions typically required for GPCR activation (e.g., ligand binding, phosphorylation, conformational

changes, etc.). Such GPCRs would be useful in screens to identify GPCR modulators, among other uses described herein.

Directed evolution is comprised of several steps. The first step is to establish a library of variants for the gene or protein of interest. The most important

5   step is to then select for those variants that entail the activity you wish to identify. The design of the screen is essential since your screen should be selective enough to eliminate non-useful variants, but not so stringent as to eliminate all variants. The last step is then to repeat the above steps using the best variant from the previous screen. Each successive cycle, can then be tailored as necessary, such as increasing the

10   stringency of the screen, for example.

Over the years, there have been a number of methods developed to introduce mutations into macromolecules. Some of these methods include, random mutagenesis, "error-prone" PCR, chemical mutagenesis, site-directed mutagenesis, and other methods well known in the art (for a comprehensive listing of current

15   mutagenesis methods, see Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982)). Typically, such methods have been used, for example, as tools for identifying the core functional region(s) of a protein or the function of specific domains of a protein (if a multi-domain protein). However, such methods have more recently been applied to the identification of macromolecule

20   variants with specific or enhanced characteristics.

Random mutagenesis has been the most widely recognized method to date. Typically, this has been carried out either through the use of "error-prone" PCR (as described in Moore, J., et al, Nature Biotechnology 14:458, (1996), or through the application of randomized synthetic oligonucleotides corresponding to specific

25   regions of interest (as described by Derbyshire, K.M. et al, Gene, 46:145-152, (1986), and Hill, DE, et al, Methods Enzymol., 55:559-568, (1987). Both approaches have limits to the level of mutagenesis that can be obtained. However, either approach enables the investigator to effectively control the rate of mutagenesis. This is particularly important considering the fact that mutations beneficial to the activity of

30   the enzyme are fairly rare. In fact, using too high a level of mutagenesis may counter or inhibit the desired benefit of a useful mutation.

While both of the aforementioned methods are effective for creating randomized pools of macromolecule variants, a third method, termed "DNA Shuffling", or "sexual PCR" (WPC, Stemmer, PNAS, 91:10747, (1994)) has recently been elucidated. DNA shuffling has also been referred to as "directed molecular

5   evolution", "exon-shuffling", "directed enzyme evolution", "in vitro evolution", and "artificial evolution". Such reference terms are known in the art and are encompassed by the invention. This new, preferred, method apparently overcomes the limitations of the previous methods in that it not only propagates positive traits, but simultaneously eliminates negative traits in the resulting progeny.

10   DNA shuffling accomplishes this task by combining the principal of in vitro recombination, along with the method of "error-prone" PCR. In effect, you begin with a randomly digested pool of small fragments of your gene, created by Dnase I digestion, and then introduce said random fragments into an "error-prone" PCR assembly reaction. During the PCR reaction, the randomly sized DNA

15   fragments not only hybridize to their cognate strand, but also may hybridize to other DNA fragments corresponding to different regions of the polynucleotide of interest – regions not typically accessible via hybridization of the entire polynucleotide. Moreover, since the PCR assembly reaction utilizes "error-prone" PCR reaction conditions, random mutations are introduced during the DNA synthesis step of the

20   PCR reaction for all of the fragments -further diversifying the potential hybridization sites during the annealing step of the reaction.

A variety of reaction conditions could be utilized to carry-out the DNA shuffling reaction. However, specific reaction conditions for DNA shuffling are provided, for example, in PNAS, 91:10747, (1994). Briefly, the DNA substrate to be

25   subjected to the DNA shuffling reaction is prepared. Preparation may be in the form of simply purifying the DNA from contaminating cellular material, chemicals, buffers, oligonucleotide primers, deoxynucleotides, RNAs, etc., and may entail the use of DNA purification kits as those provided by Qiagen, Inc., or by the Promega, Corp., for example.

30   Once the DNA substrate has been purified, it would be subjected to Dnase I digestion. About 2-4μg of the DNA substrate(s) would be digested with

110

0.0015 units of Dnase I (Sigma) per microliter in 100μl of 50mM Tris-HCL, pH 7.4/1mM MgCl2 for 10-20 min. at room temperature. The resulting fragments of 10-50bp could then be purified by running them through a 2% low-melting point agarose gel by electrophoresis onto DE81 ion-exchange paper (Whatmann) or could be

5   purified using Microcon concentrators (Amicon) of the appropriate molecular weight cutoff, or could use oligonucleotide purification columns (Qiagen), in addition to other methods known in the art. If using DE81 ion-exchange paper, the 10-50bp fragments could be eluted from said paper using 1M NaCl, followed by ethanol precipitation.

10   The resulting purified fragments would then be subjected to a PCR assembly reaction by re-suspension in a PCR mixture containing: 2mM of each dNTP, 2.2mM MgCl2, 50 mM KCl, 10mM Tris•HCL, pH 9.0, and 0.1% Triton X-100, at a final fragment concentration of 10-30ng/μl. No primers are added at this point. *Taq* DNA polymerase (Promega) would be used at 2.5 units per 100ul of

15   reaction mixture. A PCR program of 94 C for 60s; 94 C for 30s, 50-55 C for 30s, and 72 C for 30s using 30-45 cycles, followed by 72 C for 5min using an MJ Research (Cambridge, MA) PTC-150 thermocycler. After the assembly reaction is completed, a 1:40 dilution of the resulting primerless product would then be introduced into a PCR mixture (using the same buffer mixture used for the assembly reaction)

20   containing 0.8um of each primer and subjecting this mixture to 15 cycles of PCR (using 94 C for 30s, 50 C for 30s, and 72 C for 30s). The referred primers would be primers corresponding to the nucleic acid sequences of the polynucleotide(s) utilized in the shuffling reaction. Said primers could consist of modified nucleic acid base pairs using methods known in the art and referred to else where herein, or could

25   contain additional sequences (i.e., for adding restriction sites, mutating specific base-pairs, etc.).

         The resulting shuffled, assembled, and amplified product can be purified using methods well known in the art (e.g., Qiagen PCR purification kits) and then subsequently cloned using appropriate restriction enzymes.

30   Although a number of variations of DNA shuffling have been published to date, such variations would be obvious to the skilled artisan and are

encompassed by the invention. The DNA shuffling method can also be tailored to the desired level of mutagenesis using the methods described by Zhao, et al. (Nucl Acid Res., 25(6):1307-1308, (1997).

5   As described above, once the randomized pool has been created, it can then be subjected to a specific screen to identify the variant possessing the desired characteristic(s). Once the variant has been identified, DNA corresponding to the variant could then be used as the DNA substrate for initiating another round of DNA shuffling. This cycle of shuffling, selecting the optimized variant of interest, and then re-shuffling, can be repeated until the ultimate variant is obtained. Examples of

10   model screens applied to identify variants created using DNA shuffling technology may be found in the following publications: J. C., Moore, et al., J. Mol. Biol., 272:336-347, (1997), F.R., Cross, et al., Mol. Cell. Biol., 18:2923-2931, (1998), and A. Crameri., et al., Nat. Biotech., 15:436-438, (1997).

DNA shuffling has several advantages. First, it makes use of

15   beneficial mutations. When combined with screening, DNA shuffling allows the discovery of the best mutational combinations and does not assume that the best combination contains all the mutations in a population. Secondly, recombination occurs simultaneously with point mutagenesis. An effect of forcing DNA polymerase to synthesize full-length genes from the small fragment DNA pool is a background

20   mutagenesis rate. In combination with a stringent selection method, enzymatic activity has been evolved up to 16,000 fold increase over the wild-type form of the enzyme. In essence, the background mutagenesis yielded the genetic variability on which recombination acted to enhance the activity.

A third feature of recombination is that it can be used to remove

25   deleterious mutations. As discussed above, during the process of the randomization, for every one beneficial mutation, there may be at least one or more neutral or inhibitory mutations. Such mutations can be removed by including in the assembly reaction an excess of the wild-type random-size fragments, in addition to the random-size fragments of the selected mutant from the previous selection. During the next

30   selection, some of the most active variants of the polynucleotide/polypeptide/enzyme, should have lost the inhibitory mutations.

112

Finally, recombination enables parallel processing. This represents a significant advantage since there are likely multiple characteristics that would make a protein more desirable (e.g. solubility, activity, etc.). Since it is increasingly difficult to screen for more than one desirable trait at a time, other methods of molecular

5   evolution tend to be inhibitory. However, using recombination, it would be possible to combine the randomized fragments of the best representative variants for the various traits, and then select for multiple properties at once.

DNA shuffling can also be applied to the polynucleotides and polypeptides of the present invention to decrease their immunogenicity in a specified

10   host. For example, a particular variant of the present invention may be created and isolated using DNA shuffling technology. Such a variant may have all of the desired characteristics, though may be highly immunogenic in a host due to its novel intrinsic structure.. Specifically, the desired characteristic may cause the polypeptide to have a non-native structure which could no longer be recognized as a "self" molecule, but

15   rather as a "foreign", and thus activate a host immune response directed against the novel variant. Such a limitation can be overcome, for example, by including a copy of the gene sequence for a xenobiotic ortholog of the native protein in with the gene sequence of the novel variant gene in one or more cycles of DNA shuffling. The molar ratio of the ortholog and novel variant DNAs could be varied accordingly.

20   Ideally, the resulting hybrid variant identified would contain at least some of the coding sequence which enabled the xenobiotic protein to evade the host immune system, and additionally, the coding sequence of the original novel variant that provided the desired characteristics.

Likewise, the invention encompasses the application of DNA shuffling

25   technology to the evolution of polynucleotides and polypeptides of the invention, wherein one or more cycles of DNA shuffling include, in addition to the gene template DNA, oligonucleotides coding for known allelic sequences, optimized codon sequences, known variant sequences, known polynucleotide polymorphism sequences, known ortholog sequences, known homologue sequences, additional

30   homologous sequences, additional non-homologous sequences, sequences from another species, and any number and combination of the above.

In addition to the described methods above, there are a number of related methods that may also be applicable, or desirable in certain cases. Representative among these are the methods discussed in PCT applications WO 98/31700, and WO 98/32845, which are hereby incorporated by reference.

5   Furthermore, related methods can also be applied to the polynucleotide sequences of the present invention in order to evolve invention for creating ideal variants for use in gene therapy, protein engineering, evolution of whole cells containing the variant, or in the evolution of entire enzyme pathways containing polynucleotides of the invention as described in PCT applications WO 98/13485, WO 98/13487, WO

10   98/27230, WO 98/31837, and Crameri, A., et al., Nat. Biotech., 15:436-438, (1997), respectively.

Additional methods of applying "DNA Shuffling" technology to the polynucleotides and polypeptides of the present invention, including their proposed applications, may be found in US Patent No. 5,605,793; PCT Application No. WO

15   95/22625; PCT Application No. WO 97/20078; PCT Application No. WO 97/35966; and PCT Application No. WO 98/42832; PCT Application No. WO 00/09727 specifically provides methods for applying DNA shuffling to the identification of herbicide selective crops which could be applied to the polynucleotides and polypeptides of the present invention; additionally, PCT Application No. WO

20   00/12680 provides methods and compositions for generating, modifying, adapting, and optimizing polynucleotide sequences that confer detectable phenotypic properties on plant species; each of the above are hereby incorporated in their entirety herein for all purposes.

25   EXAMPLE 12
METHOD OF DISCOVERING ADDITIONAL SINGLE NUCLEOTIDE
POLYMORPHISMS (SNPS) OF THE PRESENT INVENTION

Additional SNPs may be discovered in the polynucleotides of the present invention based on comparative DNA sequencing of PCR products derived

30   from genomic DNA from multiple individuals. The genomic DNA samples may be purchased from Coriell Institute (Collingswood, NJ). PCR amplicons may be designed to cover the entire coding region of the exons using the Primer3 program.

(Rozen S 2000). Exon-intron structure of candidate genes and intron sequences may be obtained by blastn search of Genbank cDNA sequences against the human genome draft sequences. The sizes of these PCR amplicons will vary according to the exon-intron structure. All the samples may be amplified from genomic DNA (20 ng) in

5   reactions (50 μl) containing 10 mM Tris-Cl pH 8.3, 50 mM KCl, 2.5 mM MgCl₂, 150 uM dNTPs, 3 uM PCR primers, and 3.75 U TaqGold DNA polymerase (PE Biosystems).

PCR is performed in MJ Research Tetrad machines under a cycling condition of 94 degrees 10 min, 30 cycles of 94 degrees 30 sec, 60 degrees 30sec, and

10   72 degrees 30 sec, followed by 72 degrees 7 min. PCR products may be purified using QIAquick PCR purification kit (Qiagen), and may be sequenced by the dye-terminator method using PRISM 3700 automated DNA sequencer (Applied Biosystems, Foster City, CA) following the manufacturer's instruction outlined in the Owner's Manual (which is hereby incorporated herein by reference in its entirety).

15   Sequencing results may be analyzed for the presence of polymorphisms using PolyPhred software(Nickerson DA 1997; Rieder MJ 1999). All the sequence traces of potential polymorphisms may be visually inspected to confirm the presence of SNPs.

Alternative methods for identifying SNPs of the present invention are

20   known in the art. One such method involves resequencing of target sequences from individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated arrays. The strategy and principles for the design and use of such arrays are generally described in WO 95/11995.

A typical probe array used in such an analysis would have two groups

25   of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarily with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference

30   sequence, when the probe and reference sequence are aligned to maximize complementarily between the two. For each probe in the first set, there are three

corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each nucleotide in the reference sequence. The probes from the three additional probe sets would be identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position

5   in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes may be nucleotides long. Arrays tiled for multiple different references sequences may be included on the same substrate.

Publicly available sequences for a given gene can be assembled into

10   Gap4 (http:/ /www .biozentrum. unibas.ch/-biocomp/staden/Overview .html). PCR primers covering each exon, could be designed, for example, using Primer 3 (httP://www-genome.wi.mit.edu/cgi- bin/primer/primer3.cgi). Primers would not be designed in regions where there are sequence discrepancies between reads. Genomic DNA could be amplified from at least two individuals using 2.5 pmol each primer, 1.5

15   mM MgCl2, 100 ~M dNTPs, 0.75 ~M AmpliTaq GOLD polymerase, and about 19ng DNA in a 15 ul reaction. Reactions could be assembled using a PACKARD MultiPROBE robotic pipetting station and then put in MJ 96-well tetrad thermocyclers (96°C for minutes, followed by cycles of 96°C for seconds, 59°C for 2 minutes, and 72°C for 2 minutes). A subset of the PCR assays for each individual

20   could then be run on 3% NuSieve gels in 0.5X TBE to confirm that the reaction worked.

For a given DNA, 5ul (about 50 ng) of each PCR or RT -PCR product could be pooled (Final volume = 150-200 ul). The products can be purified using QiaQuick PCR purification from Qiagen. The samples would then be eluted once in

25   35ul sterile water and 4 ul lOX One-Phor-All buffer (Pharmacia). The pooled samples are then digested with 0.2u DNaseI (Promega) for 10 minutes at 37°C and then labeled with 0.5 nmols biotin-N6- ddATP and 15u Terminal Transferase (GibcoBRL Life Technology) for 60 minutes at 37°C. Both fragmentation and labeling reactions could be terminated by incubating the pooled sample for 15 minutes

30   at 100°C.

Low-density DNA chips (Affymetrix,CA) may be hybridized following the manufacturer's instructions. Briefly, the hybridization cocktail

consisted of 3M TMACI, mM Tris pH 7.8, 0.01% Triton X-100, 100 mg/ml herring

sperm DNA {Gibco BRL), 200 pM control biotin-labeled oligo. The processed PCR

products are then denatured for 7 minutes at 100°C and then added to prewarmed

{37°C) hybridization solution. The chips are hybridized overnight at 44°C. Chips are

5   ished in 1X SSPET and 6X SSPET followed by staining with 2 ug/ml SARPE and 0.5

mg/ml acetylated BSA in 200 ul of 6X SSPET for 8 minutes at room temperature.

Chips are scanned using a Molecular Dynamics scanner.

Chip image files may be analyzed using Ulysses {Affymetrix, CA)

which uses four algorithms to identify potential polymorphisms. Candidate

10   polymorphisms may be visually inspected and assigned a confidence value: where

high confidence candidates display all three genotypes, while likely candidates show

only two genotypes {homozygous for reference sequence and heterozygous for

reference and variant). Some of the candidate polymorphisms may be confirmed by

ABI sequencing. Identified polymorphisms could then be compared to several

15   databases to determine if they are novel.

EXAMPLE 13

METHOD OF DETERMINING THE ALLELE FREQUENCY FOR EACH SNP OF

THE PRESENT INVENTION.

Allele frequencies of these polymorphisms may be determined by

20   genotyping various DNA samples (Coriell Institute; Collingswood, NJ) using FP-TDI

assay (Chen X 1999). Automated genotyping calls may be made with an allele

calling software developed by Joel Hirschorn (Whitehead Institute/MIT Center for

Genome Research, personal communication).

Briefly, the no template controls (NTCs) may be labeled accordingly in

25   column C. The appropriate cells may be completed in column L indicating whether

REF (homozygous ROX) or VAR (homozygous TAMRA) are expected to be rare

genotypes (<10% of all samples) – the latter is important in helping the program to

identify rare homozygotes. The number of 96 well plates genotyped in cell P2 are

noted (generally between 0.5 and 4) - the program works best if this is accurate. No

30   more than 384 samples can be analyzed at a time. The pairs of mP values from the

LJL may be pasted into columns E and F; making sure there may be no residual data

is left at the bottom fewer than 384 data points are provided. The DNA names may be

provided in columns A, B or C; column I will be a concatenation of columns A, B and C. In addition, the well numbers for each sample may be also provided in column D.

With the above information provided, the program should automatically cluster the points and identify genotypes. The program works by
5   converting the mP values into polar coordinates (distance from origin and angle from origin) with the angle being on a scale from 0 to 2; heterozygotes are placed as close to 1 as possible.

The cutoff values in columns L and M may be adjusted as desired.

Expert parameters: The most important parameters are the maximum
10   angle for REF and minimum angle for VAR. These parameters may need to be changed in a particularly skewed assay which may be observed when an REF or VAR cluster is close to an angle of 1 and has called as a failed or HETs.

Other parameters are low and high cutoffs that are used to determine which points are considered for the determination of edges of the clusters. With small
15   numbers of data points, the high cutoff may need to be increased (to 500 or so). This may be the right thing to do for every assay, but certainly when the program fails to identify a small cluster with high signal.

NTC TAMRA and ROX indicate the position of the no template control or failed samples as estimated by the computer algorithm.
20   No signal = mP< is the threshold below which points are automatically considered failures. "Throw out points with signal above" is the TAMRA or ROX mP value above which points are considered failures. The latter may occasionally need to be adjusted from 250 to 300, but caveat emptor for assays with signals >250. 'Lump' or 'split' describes a subtle difference in the way points are grouped into
25   clusters. Lump generally is better. 'HETs expected' in the rare case where only homozygotes of either class are expected (e.g. a study of X chromosome SNPs in males), change this to "N".

Notes on method of clustering: The origin is defined by the NTCs or other low signal points (the position of the origin is shown as "NTC TAMRA" and
30   "NTC ROX"); the points with very low or high signal are not considered initially. The program finds the point farthest from the origin and calls that a HET; the ROX/TAMRA ratio is calculated from this point, placing the heterozygotes at 45

degrees from the origin (an angle of "1"). The angles from the origin are calculated (the scale ranges from 0 to 2) and used to define clusters. A histogram of angles is generated. The cluster boundaries are defined by an algorithm that takes into account the shape of the histogram. The homozygote clusters are defined as the leftmost and

5   rightmost big clusters (unless the allele is specified as being rare, in which case the cluster need not be big). The heterozygote is the biggest cluster in between the REF and VAR. If there are two equal clusters, the one best-separated from REF and VAR is called HET. All other clusters are failed. Some fine tuning is applied to lump in scattered points on the edges of the clusters (if "Lump" is selected). The boundaries

10   of the clusters are "Angles" in column L.

Once the clusters are defined, the interquartile distance of signal intensity is defined for each cluster. Points falling more than 3 or 4 interquartiles from the mean are excluded. (These are the "Signal cutoffs" in column M).

Allele frequency of the B1 receptor R317Q variant (AE103s1) is as

15   follows. 7% in African Americans (7/94), 0% in Caucasians (0/94), 0% in Asians (0/60), and 0% in Amerindians (0/20). Higher frequency of this form in African Americans than in Caucasians matches the profile of a potential genetic risk factor for angioedema, which is observed more frequently in African Americans than in Caucasians (Brown NJ 1996; Brown NJ 1998; Agostoni A 1999; Coats 2000).

20   The invention encompasses additional methods of determinig the allelic frequency of the SNPs of the present invention. Such methods may be known in the art, some of which are described elsewhere herein.

EXAMPLE 14

ALTERNATIVE METHODS OF DETECTING POLYMORPHISMS

25   ENCOMPASSED BY THE PRESENT INVENTION.

Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include

30   whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in

which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. See generally PCR

5   Technology: Principles and Applications for DNA Amplification (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); PCR Protocols: A Guide to Methods and Applications (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., Nucleic Acids Res. 19, 4967 (1991); Eckert et al., PCR Methods and Applications 1, (1991); PCR (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202.

10   Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, Genomics 4:560 (1989), Landegren et al., Science 241:1077 (1988), transcription amplification (Kwoh et al., Proc. Natl. Acad. Sci. USA 86, 1173 (1989), and self-sustained sequence replication (Guatelli et al., Proc. Nat. Acad. Sci. USA, 87:1874 (1990)) and nucleic acid based sequence amplification

15   (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

Additional methods of amplification are known in the art or are

20   described elsewhere herein.

Detection of Polymorphisms in Target DNA

There are two distinct types of analysis of target DNA for detecting polymorphisms. The first type of analysis, sometimes referred to as de novo characterization, is carried out to identify polymorphic sites not previously

25   characterized (i.e., to identify new polymorphisms). This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified,

30   and the frequencies of such alleles/haplotypes in the population can be determined. Additional allelic frequencies can be determined for subpopulations characterized by

criteria such as geography, race, or gender. The de novo identification
of polymorphisms of the invention is described in the Examples section.

The second type of analysis determines which form(s) of a
characterized (known) polymorphism are present in individuals under test. Additional
5   methods of analysis are known in the art or are described elsewhere herein.

Allele-Specific Probes

The design and use of allele-specific probes for analyzing
polymorphisms is described by e.g., Saiki et al., Nature 324,163-166 (1986);
Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed
10   that hybridize to a segment of target DNA from one individual but do not hybridize to
the corresponding segment from another individual due to the presence of different
polymorphic forms in the respective segments from the two individuals.
Hybridization conditions should be sufficiently stringent that there is a significant
difference in hybridization intensity between alleles, and preferably an essentially
15   binary response, whereby a probe hybridizes to only one of the alleles. Some probes
are designed to hybridize to a segment of target DNA such that the polymorphic site
aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either
the 8 or 9 position) of the probe. This design of probe achieves good discrimination
in hybridization between different allelic forms.
20   Allele-specific probes are often used in pairs, one member of a pair
showing a perfect match to a reference form of a target sequence and the other
member showing a perfect match to a variant form. Several pairs of probes can then
be immobilized on the same support for simultaneous analysis of multiple
polymorphisms within the same target sequence.

25   Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic
acid arrays, some examples of which are described in WO 95/11995. The same arrays
or different arrays can be used for analysis of characterized polymorphisms. WO
95/11995 also describes sub arrays that are optimized for detection of a variant form
30   of a precharacterized polymorphism. Such a subarray contains probes designed to be
complementary to a second reference sequence, which is an allelic variant of the first

reference sequence. The second group of probes is designed by the same principles as described, except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particularly useful for analyzing short subsequences of the primary reference sequence in which

5   multiple mutations are expected to occur within a short distance commensurate with the length of the probes (e.g., two or more mutations within 9 to bases).

Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to

10   which the primer exhibits perfect complementarity. See Gibbs, Nucleic Acid Res. 17,2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base

15   mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing elongation from the primer (see, e.g., WO

20   93/22456).

Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam - Gilbert method (see Sambrook et al., Molecular Cloning, A Laboratory

25   Manual (2nd Ed., CSHP, New York 1989); Zyskind et al., Recombinant DNA Laboratory Manual, (Acad. Press, 1988)).

Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different

30   alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., PCR Technology.

Principles and Applications for DNA Amplification, (W .H. Freeman and Co, New

York, 1992), Chapter 7.

Single-Strand Conformation Polymorphism Analysis

        Alleles of target sequences can be differentiated using single-strand

5   conformation polymorphism analysis, which identifies base differences by alteration

in electrophoretic migration of single stranded PCR products, as described in Orita et

al., Proc. Nat. Acad. Sci. 86,2766-2770 (1989). Amplified PCR products can be

generated as described above, and heated or otherwise denatured, to form single

stranded amplification products. Single-stranded nucleic acids may refold or form

10 secondary structures which are partially dependent on the base sequence. The

different electrophoretic mobilities of single-stranded amplification products can be

related to base-sequence differences between alleles of target sequences.

Single Base Extension

        An alternative method for identifying and analyzing polymorphisms is

15 based on single-base extension (SBE) of a fluorescently-labeled primer coupled with

fluorescence resonance energy transfer (FRET) between the label of the added base

and the label of the primer. Typically, the method, such as that described by Chen et

al., (PNAS 94:10756-61 (1997), uses a locus-specific oligonucleotide primer labeled

on the 5' terminus with 5-carboxyfluorescein (F AM). This labeled primer is designed

20 so that the 3' end is immediately adjacent to the polymorphic site of interest. The

labeled primer is hybridized to the locus, and single base extension of the labeled

primer is performed with fluorescently-labeled dideoxyribonucleotides (ddNTPs) in

dye-terminator sequencing fashion. An increase in fluorescence of the added ddNTP

in response to excitation at the wavelength of the labeled primer is used to infer the

25 identity of the added nucleotide.

        The contents of all patents, patent applications, published PCT

applications and articles, books, references, reference manuals and abstracts cited

herein are hereby incorporated by reference in their entirety to more fully describe the

30 state of the art to which the invention pertains.

As various changes can be made in the above-described subject matter without departing from the scope and spirit of the present invention, it is intended that all subject matter contained in the above description, or defined in the appended claims, be interpreted as descriptive and illustrative of the present invention. Many

5   modifications and variations of the present invention are possible in light of the above teachings.

REFERENCES

1. Rees, S., Brown, S., Stables, J.: "Reporter gene systems for the study of G Protein Coupled Receptor signalling in mammalian cells". In Milligan G. (ed.): Signal Transduction: A practical approach. Oxford: Oxford University Press, 1999: 171-221.

2. Alam, J., Cook, J.L.: "Reporter Genes: Application to the study of mammalian gene transcription". Anal. Biochem. 1990; 188: 245-254.

3. Selbie, L.A. and Hill, S.J.: "G protein-coupled receptor cross-talk: The fine-tuning of multiple receptor-signaling pathways". TiPs. 1998; 19: 87-93.

4. Boss, V., Talpade, D.J., and Murphy, T.J.: "Induction of NFAT mediated transcription by Gq-coupled Receptors in lympoid and non-lymphoid cells". JBC. 1996; 271: 10429-10432.

5. George, S.E., Bungay, B.J., and Naylor, L.H.: "Functional coupling of endogenous serotonin (5-HT1B) and calcitonin (C1a) receptors in CHO cells to a cyclic AMP-responsive luciferase reporter gene". J. Neurochem. 1997; 69: 1278-1285.

6. Suto, CM, Igna DM: "Selection of an optimal reporter for cell-based high throughput screening assays". J. Biomol. Screening. 1997; 2: 7-12.

7. Zlokarnik, G., Negulescu, P.A., Knapp, T.E., More, L., Burres, N., Feng, L., Whitney, M., Roemer, K., and Tsien, R. Y. "Quantitation of transcription and clonal selection of single living cells with a B-Lactamase Reporter". Science. 1998; 279: 84-88.

8. S. Fiering et. al., Genes Dev. 4, 1823 (1990).

9. J. Karttunen and N. Shastri, PNAS 88, 3972 (1991).

10. Hawes, B. E., Luttrell. L.M., van Biesen, T., and Lefkowitz, R.J. (1996) JBC 271, 12133-12136.

11. Gilman, A.G. (1987) Annul. Rev. Biochem. 56, 615-649.

12. Maniatis et al., Cold Spring Harbor Press, 1989.

13. Salcedo, R., Ponce, M.L., Young, H.A., Wasserman, K., Ward, J.M., Kleinman, H.K., Oppenheim, J.J., Murphy, W.J. "Human endothelial cells express CCF2 and respond to MCP-1: direct role of MCP-1 in angiogenesis and tumor progression". Blood. 2000; 96 (1): 34-40.

14. Sica, A., Saccani, A., Bottazzi, B., Bernasconi, S., Allavena, P., Gaetano, B., LaRossa, G., Scotton, C., Balkwill F., Mantovani, A. "Defective expression of the monocyte chemotactic protein 1 receptor CCF2 in macrophages associated with human ovarian carcinoma". J. Immunology. 2000; 164: 733-8.

15. Kypson, A., Hendrickson, S., Akhter, S., Wilson, K., McDonald, P., Lilly, R., Dolber, P., Glower, D., Lefkowitz, R., Koch, W. "Adenovirus-mediated gene transfer of the B2 AR to donor hearts enhances cardiac function". Gene Therapy. 1999; 6: 1298-304.

16. Dorn, G.W., Tepe, N.M., Lorenz, J.N., Kock, W.J., Ligget, S.B. "Low and high level transgenic expression of B2AR differentially affect cardiac hypertrophy and function in Galpha q-overexpressing mice". PNAS. 1999; 96: 6400-5.

17. J. Wess. "G protein coupled receptor: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition". FASEB. 1997; 11:346-354.

18. Whitney, M, Rockenstein, E, Cantin, G., Knapp, T., Zlokarnik, G., Sanders,

P., Durick, K., Craig, F.F., and Negulescu, P.A. "A genome-wide functional assay of signal transduction in living mammalian cells". Nature Biotech. 1998; 16: 1329-1333.

19. BD Biosciences: FACS Vantage SE Training Manual. Part Number 11-11020-00 Rev. A. August 1999.

20. Chen, G., Jaywickreme, C., Way, J., Armour S., Queen K., Watson., C., Ignar, D., Chen, W.J., Kenakin, T. "Constitutive Receptor systems for drug discovery". J. Pharmacol. Toxicol. Methods 1999; 42: 199-206.

21. Blahos, J., Fischer,T., Brabet, I., Stauffer,D., Rovelli, G., Bockaert, J., and Pin, J.-P. "A novel Site on the G alpha-protein that Rocognized Heptahelical Receptors". J.Biol. Chem. 2001; 275, No. 5, 3262-69.

22. Offermanns, S. & Simon, M.I. "G alpha 15 and G alpha 16 Couple a Wide Variety of Receptors to Phospholipase C". J. Biol. Chem. 1995; 270, No. 25, 15175-80.

WHAT IS CLAIMED IS:

1.   An isolated polynucleotide selected from the group consisting of:

        (a)   an isolated polynucleotide encoding a human G-protein coupled
                receptor, or functional fragment thereof, comprising the amino acid
5   sequence as set forth in SEQ ID NO:2;

        (b)   An isolated composition comprising the polynucleotide according to
                (a).

        (c)   An isolated polynucleotide comprising SEQ ID NO:1;

        (d)   An isolated polynucleotide having the nucleic acid sequence of ATCC
10   Accession No. PTA-2966;

        (e)   An isolated polynucleotide having the nucleic acid sequence according
                to nucleotides 4 to 1524 of SEQ ID NO:1, wherein said nucleotides
                encode a polypeptide of SEQ ID NO:2 minus the start codon;

        (f)   An isolated polynucleotide having the nucleic acid sequence according
15   to nucleotides 1 to 1524 of SEQ ID NO:1, wherein said nucleotides
                encode a polypeptide of SEQ ID NO:2 including the start codon;

        (g)   A polynucleotide which is fully complementary to the polynucleotide
                according to (a) thru (f); and

        (h)   A hybridization probe comprising the polynucleotide according to (a)
20   thru (g).

2.   An expression vector containing the polynucleotide according to claim 1.

3.   A host cell containing the expression vector according to claim 2.

4.   A substantially purified G-protein coupled receptor polypeptide selected from
the group consisting of:

25   (a)   A substantially purified G-protein coupled receptor polypeptide,
                comprising an amino acid sequence as set forth in SEQ ID NO:2.

        (b)   The polypeptide according to (a), wherein the amino acid sequence
                differs from SEQ ID NO:2 only by conservative substitutions;

        (c)   An isolated and substantially purified G-protein coupled receptor
30   polypeptide encoded by the nucleic acid sequence of ATCC Accession
                No. PTA-2966;

127

(d)   An isolated polypeptide having the amino acid sequence according to
        amino acids 2 to 508 of SEQ ID NO:2, wherein said amino acid
        encode a polypeptide of SEQ ID NO:2 minus the start methionine;

(e)   An isolated polypeptide having the amino acid sequence according to
        amino acids 1 to 508 of SEQ ID NO:2, wherein said amino acid
        encode a polypeptide of SEQ ID NO:2 including the start methionine;

(f)   A substantially purified fragment of the G-protein coupled receptor
        polypeptide according to any one of (a) to (e).

5.   A substantially purified fusion protein comprising an amino acid sequence as
set forth in SEQ ID NO:2 and an amino acid sequence of an Fc portion of a human
immunoglobulin protein.

6.   A pharmaceutical composition comprising the polypeptide, or a functional
fragment thereof, according to any one of claims 1, 2, 3, 4, or 5, and a
pharmaceutically acceptable diluent or excipient.

7.   A purified antibody which binds specifically to the polypeptide according to
claim 4, or an antigenic epitope thereof.

8.   A method of screening a library of molecules or compounds with a
polynucleotide to identify at least one molecule or compound therein which
specifically binds to the polynucleotide sequence, comprising:

(a)   combining the polynucleotide according to claim 1, with a library of
        molecules or compounds under conditions to allow specific binding;
        and

(b)   detecting specific binding, thereby identifying a molecule or
        compound, which specifically binds to a G-protein coupled receptor-
        encoding polynucleotide sequence.

9.   The method according to claim 8, wherein the candidate compounds are small
molecules, therapeutics, biological agents, or drugs.

10.   A method of screening for candidate compounds capable of modulating
activity of a G-protein coupled receptor-encoding polypeptide, comprising:

(a)   contacting a test compound with a cell or tissue expressing the
        polypeptide according to claim 4; and

128

(b)   selecting as candidate modulating compounds those test compounds
       that modulate activity of the G-protein coupled receptor polypeptide.

11.   A method of treating a neurological disorder in a mammal comprising
administration of the G-protein coupled receptor polypeptide or homologue according
to any one of claims 19-21, or claim 23 or 24, in an amount effective to treat the
neurological disorder.

12.   A substantially purified G-protein coupled receptor polypeptide consisting of
an amino acid sequence as set forth in SEQ ID NO:2.

13.   The polypeptide according to claim 12, wherein the amino acid sequence
differs from SEQ ID NO:2 only by conservative substitutions.

14.   An isolated and purified polynucleotide encoding a human G-protein coupled
receptor, or functional fragment thereof, consisting of the amino acid sequence as set
forth in SEQ ID NO:2.

15.   A method of treating a disease, disorder, or condition related to the brain
comprising administering the G-protein coupled receptor polypeptide or homologue
according to claim 12 or 13 in an amount effective to treat the brain-related disorder.

16.   The polypeptide of claim 12 or 13, further comprising the polypeptide
expressed in the caudate nucleus, substantia nigra, thalamus, amygdala, hippocampus,
cerebellum, and corpus collosum.

17.   A cell comprising NFAT/CRE and the polypeptide of claim 12 or 13.

18.   A cell comprising NFAT G alpha 15 and the polypeptide of claim 12 or 13.

19.   A method of screening for candidate compounds capable of modulating
activity of a G-protein coupled receptor-encoding polypeptide, comprising:

(a)   contacting a test compound with a cell or tissue expressing the
       polypeptide according to claim 12 or 13; and

(b)   selecting as candidate modulating compounds those test compounds
       that modulate activity of the G-protein coupled receptor polypeptide.

20.   The method according to claim 19, wherein the candidate compounds are
agonists or antagonists of G-protein coupled receptor activity.

21.   The method according to claim 20, wherein the polypeptide activity is
associated with the brain.

22.   The method according to claim 20, wherein the candidate modulating compounds are peptides.

FIG. 1

ATGACGTCCACCTGCACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCC
CTCTCCAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGGTTATC
TTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTGTTGCAGCGCAAGCCG
CAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAACCTCCTCGTCACCGACCTGCTGCAG
ATTTCGCTCGTGGCCCCCTGGGTGGTGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAAC
AGCCACTTCTGCACGGCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAAC
ACCATTGTCTTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCTACCCG
TCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACCTGGATTGTGGCCATC
CTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCAGGCTGCCTTTGATGAGCGCAATGCT
CTCTGCTCCATGATCTGGGGGGCCAGCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTC
ATCGTCATTCCACTGATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGG
AGGCAGCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCAAGGAC
TGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAGTTCCAGGATGAGAGT
GAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGCCAAGGAGGGCAGAATGGAAGCCAAG
GACGGCAGCCTGAAGGCCAAGGAAGGAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCC
AGGGGCAGCGAGGAGGTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGT
AAGGAAGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCACAGAG
GTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTTGGTGAAGACGACATC
AATTTCAGTGAGGATGACGTCGAGGCAGTGAACATCCCGGAGAGCCTCCCACCCAGTCGT
CGTAACAGCAACAGCAACCCTCCTCTGCCCAGGTGCTACCAGTGCAAAGCTGCTAAAGTG
ATCTTCATCATCATTTTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTC
CTGGCCGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAATCATC
TGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGCTACATGCACAAGACC
ATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTTCTTCTGCAAGGAAAAGCCCCCGAAA
GAAGATAGCCACCCAGACCTGCCCGGAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCT
TCCTACGATTCTGCTACTTTTCCTTGA

FIG.2

```
MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP
QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN
TIVLVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA
LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD
CVENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEA
RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDI
NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYCFLAV
LAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK
EDSHPDLPGTEGGTEGKIVPSYDSATFP
```

FIG. 3

```
GCAACCTGTCTCACGCCCTCTGGCTGTTGCC
```

FIG. 4

AGTTAGTTCTAAGGCAAACCTT

FIG. 5

```
  1  MTSTCTNSTR  ESNSSHTCMP  LSKMPISLAH  GIIRSTVLVI  FLAASFVGNI
 51  VLALVLQRKP  QLLQVTNRFI  FNLLVTDLLQ  ISLVAPWVVA  TSVPLFWPLN
101  SHFCTALVSL  THLFAFASVN  TIVLVSVDRY  LSIIHPLSYP  SKMTQRRGYL
151  LLYGTWIVAI  LQSTPPLYGW  GQAAFDERNA  LCSMIWGASP  SYTILSVVSF
201  IVIPLIVMIA  CYSVVFCAAR  RQHALLYNVK  RHSLEVRVKD  CVENEDEEGA
251  EKKEEFQDES  EFRRQHEGEV  KAKEGRMEAK  DGSLKAKEGS  TGTSESSVEA
301  RGSEEVRESS  TVASDGSMEG  KEGSTKVEEN  SMKADKGRTE  VNQCSIDLGE
351  DDMEFGEDDI  NFSEDDVEAV  NIPESLPPSR  RNSNSNPPLP  RCYQCKAAKV
401  IFIIIFSYVL  SLGPYCFLAV  LAVWVDVETQ  VPQWVITIII  WLFFLQCCIH
451  PYVYGYMHKT  IKKEIQDMLK  KFFCKEKPPK  EDSHPDLPGT  EGGTEGKIVP
501  SYDSATFP
```

FIG. 6A

```
ACM4_CHICK   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
YDBM_CAEEL   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_HUMAN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_MOUSE   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_FUGRU   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5HT_LYMST    ~MANFTFGDLALDVARMGGLASTPSGLRSTGLTTPGLSPT
A1AD_HUMAN   MTFRDLLSVSFEGPRPDSSAGGSSAGGGGGSAGGAAPSEG
A1AD_MOUSE   MTFRDILSVTFEGPRASSSTGGSGAGGGAGTVG...P.EG
Q13675       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Q13729       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O60451       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RAT     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O54913       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_BOVIN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_CANFA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_HUMAN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_ORYLA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O96716       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O75963       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~


ACM4_CHICK   ~~~~~~~~~~~~~~~MHNLSAQPWQAKMANLTYDNVTLSN
YDBM_CAEEL   MCFAEKGEGAGEDVDHHSLFC.P.KKLVGNL...KGFIRN
5H1A_HUMAN   ~~~~~~~MDV..LSPGQ...GNNTT....SPPAPFETGGN
5H1A_MOUSE   ~~~~~~~MDM..FSLGQ...GNNTT....TSLEPFGTGGN
5H1A_FUGRU   ~~~~~~~MDLRATSSND...SNATSGYSDTAAVDWDEGEN
5HT_LYMST    GLVTSDFNDSYGLT.GQFINGSHSSRSRDNASANDTSATN
A1AD_HUMAN   PAVGGVPGGAGGGGGVVGAGSGEDNRSSAGEPGSAGAGGD
A1AD_MOUSE   PAVGGVP.GATGGSAVVGTGSGEDNQSSTAEAGAA.ASGE
Q13675       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
Q13729       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
O60451       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
A1AA_RAT     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVLLSENASEG
O54913       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVLLSENASEG
A1AA_BOVIN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
A1AA_CANFA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
A1AA_RABIT   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
A1AA_HUMAN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVFLSGNASDS
A1AA_ORYLA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTPSSVTLNC
O96716       ~~~~~~~~~~~~~~~~~MSANTTVSPTETTANLTANSTEA
O75963       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSLNSSLS
HGPRBMY8     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTSTCTNSTRESNS
```

FIG. 6B

```
ACM4_CHICK    RSEVAIQPPTNYKTVELVFIATVTGSLS..LVTVVGNILV
YDBM_CAEEL    QYH.......QHETIQ.ILKGSALFLLV..LWTIFANSLV
5H1A_HUMAN    TTG...ISDVTVSYQ..VITSLLLGTLI..FCAVLGNACV
5H1A_MOUSE    DTG...LSNVTFSYQ..VITSLLLGTLI..FCAVLGNACV
5H1A_FUGRU    ATGSGSLPDPELSYQ..IITSLFLGALI..LCSIFGNSCV
5HT_LYMST     MTDDRYWSLTVYSHEHLVLTSVILGLFV..LCCIIGNCFV
A1AD_HUMAN    V.NGTAAVGGLVVSAQGVGVGVFLAAFI..LMAVAGNLLV
A1AD_MOUSE    V.NGSAAVGGLVVSAQGVGVGVFLAAFI..LTAVAGNLLV
Q13675        S.NCTQPPAP.VNISKAILLGVILGGLI..LFGVLGNILV
Q13729        S.NCTQPPAP.VNISKAILLGVILGGLI..LFGVLGNILV
O60451        S.NCTQPPAP.VNISKAILLGVILGGLI..LFGVLGNILV
A1AA_RAT      S.NCTHPPAP.VNISKAILLGVILGGLI..IFGVLGNILV
O54913        S.NCTHPPAQ.VNISKAILLGVILGGLI..IFGVLGNILV
A1AA_BOVIN    S.NCTHPPPP.VNISKAILLGVILGGLI..LFGVLGNILV
A1AA_CANFA    S.NCTHPPAP.VNISKAILLGVILGGLI..IFGVLGNILV
A1AA_RABIT    S.NCTHPPAP.VNISKAILLGVILGGLI..LFGVLGNILV
A1AA_HUMAN    S.NCTQPPAP.VNISKAILLGVILGGLI..LFGVLGNILV
A1AA_ORYLA    S.NCSHVLAPELNTVKAVVLGMVLGIFI..LFGVIGNILV
O96716        SVGSCFAPNPYSAGVQAV.LGLITVILI..LLTVIGNVLV
O75963        CRKELSNLTEEEGGEGGVIITQFIAIIVITIFVCLGNLVI
HGPRBMY8      SHTCMPLSKMPISLAHGIIRSTVL..VIFLAASFVGNIVL


ACM4_CHICK    MLSIKVNRQLQTVNNYFLFSLACADLIIGVFSMNLYTVYI
YDBM_CAEEL    FIVLYKNPRLQTVPNLLVGNLAFSDLALGLIVLPLSSVYA
5H1A_HUMAN    VAAIALERSLQNVANYLIGSLAVTDLMVSVLVLPMAALYQ
5H1A_MOUSE    VAAIALERSLQNVANYLIGSLAVTDLMVSVLVLPMAALYQ
5H1A_FUGRU    VAAIALERSLQNVANYLIGSLAVTDLMVSVLVLPMAALYQ
5HT_LYMST     IAAVMLERSLHNVANYLILSLAVADLMVAVLVMPLSVVSE
A1AD_HUMAN    ILSVACNRHLQTVTNYFIVNLAVADLLLSATVLPFSATME
A1AD_MOUSE    ILSVACNRHLQTVTNYFIVNLAVADLLLSAAVLPFSATME
Q13675        ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
Q13729        ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
O60451        ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_RAT      ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
O54913        ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_BOVIN    ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_CANFA    ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_RABIT    ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_HUMAN    ILSVACHRHLHSVTHYYIVNLAVADLLLTSTVLPFSAIFE
A1AA_ORYLA    ILSVVCHRHLQTVTYYFIVNLAVADLLLSTVLPFSAIFE
O96716        ILAVTCHRKMRTVTNFFIVSLACADLSVGITVLPFAATND
O75963        VVTLYKKSYLLTLSNKFVFSLTLSNFLLSVLVLPFVVTSS
HGPRBMY8      ALVLQRKPQLLQVTNRFIFNLLVTDLLQISLVAPWVVATS
```

FIG. 6C

```
ACM4_CHICK    IKGYWPLGAVVCDIWLALDYVVSNASVMNLLIISFDRYFC
YDBM_CAEEL    IAGEWVFPDALCEVFVSADILCSTASIWNLSIVGLDRYWA
5H1A_HUMAN    VLNKWTLGQVTCDLFIALDVLCCTSSILHLCAIALDRYWA
5H1A_MOUSE    VLNKWTLGQVTCDLFIALDVLCCTSSILHLCAIALDRYWA
5H1A_FUGRU    VLNKWTLGQDICDLFIALDVLCCTSSILHLCAIALDRYWA
5HT_LYMST     ISKVWFLHSEVCDMWISVDVLCCTASILHLVAIAMDRYWA
A1AD_HUMAN    VLGFWAFGRAFCDVWAAVDVLCCTASILSLCTISVDRYVG
A1AD_MOUSE    VLGFWPFGRTFCDVWAAVDVLCCTASILSLCTISVDRYVG
Q13675        VLGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
Q13729        VLGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
O60451        VLGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
A1AA_RAT      ILGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
O54913        ILGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
A1AA_BOVIN    ILGYWAFGRVFCNVWAAVDVLCCTASIMGLCIISIDRYIG
A1AA_CANFA    ILGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
A1AA_RABIT    ILGYWAFGRVFCNIWAAVDVLCCTASIISLCVISIDRYIG
A1AA_HUMAN    VLGYWAFGRVFCNIWAAVDVLCCTASIMGLCIISIDRYIG
A1AA_ORYLA    ILDRWVFGRVFCNIWAAVDVLCCTASIMSLCVISVDRYIG
O96716        ILGYWPFGG.YCDVWVSFDVLNSTASIILNLVVIAFDRFLA
O75963        IRREWIFGVVWCNFSALLYLLISSASMLTLGVIAIDRYYA
HGPRBMY8      VPLFWPLNSHFCTALVSLTHLFAFASVNTIVLVSVDRYLS


ACM4_CHICK    VTKPLTYPARRTTKMAGLMIAAAWILSFILWAPAI.LFW.
YDBM_CAEEL    ITSPVAYMSKRNKRTAGIMILSVWISSALISLAPL.LGWK
5H1A_HUMAN    ITDPIDYVNKRTPRRAAALISLTWLIGFLISIPPM.LGW.
5H1A_MOUSE    ITDPIDYVNKRTPRRAAALISLTWLIGFLISIPPM.LGW.
5H1A_FUGRU    ITDPIDYVNKRTPRRAAVLISVTWLIGFSISIPPM.LGW.
5HT_LYMST     VTS.IDYIRRRSARRILLMIMVVWIVALFISIPPL.FGW.
A1AD_HUMAN    VRHSLKYPAIMTERKAAAILALLWVVALVVSVGPL.LGW.
A1AD_MOUSE    VRHSLKYPAIMTERKAAAILALLWAVALVVSVGPL.LGW.
Q13675        VSYPLRYPTIVTQRRGLMALLCVWALSLVISIGPL.FGW.
Q13729        VSYPLRYPTIVTQRRGLMALLCVWALSLVISIGPL.FGW.
O60451        VSYPLRYPTIVTQRRGLMALLCVWALSLVISIGPL.FGW.
A1AA_RAT      VSYPLRYPTIVTQRRGVRALLCVWVLSLVISIGPL.FGW.
O54913        VSYPLRYPTIVTQRRGVRALLCVWALSLVISIGPL.FGW.
A1AA_BOVIN    VSYPLRYPTIVTQKRGLMALLCVWALSLVISIGPL.FGW.
A1AA_CANFA    VSYPLRYPTIVTQKRGLMALLCVWALSLVISIGPL.FGW.
A1AA_RABIT    VSYPLRYPTIVTQRRGLRALLCVWAFSLVISVGPL.FGW.
A1AA_HUMAN    VSYPLRYPTIVTQRRGLMALLCVWALSLVISIGPL.FGW.
A1AA_ORYLA    VSYPLRYPAIMTKRRALLAVMLLWVLSVIISIGPL.FGW.
O96716        ITAPFTYHTRMTERTAGILIATVWGISLVVSFLPIQAGWY
O75963        VLYPMVYPMKITGNRAVMALVYIWLHSLIGCLPPL.FGWS
HGPRBMY8      IIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPL.YGWG
```

FIG. 6D

```
ACM4_CHICK    QFIVGKRTVHE.....RECYIQFLSNPAVTFGTAIAAFYL
YDBM_CAEEL    QTAQTPNLIYEKNNTVRQC..TFLDLPSYTVYSATGSFFI
5H1A_HUMAN    R.TP.EDRSDPDA.....CTIS..KDHGYTIYSTFGAFYI
5H1A_MOUSE    R.AP.EDRSNPNE.....CTIS..KDHGYTIYSTFGAFYI
5H1A_FUGRU    R.SA.EDRANPDA.....CIIS..QDPGYTIYSTFGAFYI
5HT_LYMST     R.DPNNDPDKTGT.....CIIS..QDKGYTIFSTVGAFYL
A1AD_HUMAN    K.EPVPP.......DERFCGIT..EEAGYAVFSSVCSFYL
A1AD_MOUSE    K.EPVPP.......DERFCGIT..EEVGYAIFSSVCSFYL
Q13675        R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYL
Q13729        R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYL
O60451        R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYL
A1AA_RAT      R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYV
O54913        R.QQAPE.......DETICQIN..EEPGYVLFSALGSFYV
A1AA_BOVIN    R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYV
A1AA_CANFA    R.QPAPE.......DETICQIT..EEPGYVLFSALGSFYV
A1AA_RABIT    R.QPAPD.......DETICQIN..EEPGYVLFSALGSFYV
A1AA_HUMAN    R.QPAPE.......DETICQIN..EEPGYVLFSALGSFYL
A1AA_ORYLA    K.EPAPE.......DETVCKIT..EEPGYAIFSAVGSFYL
O96716        R.DNQSEEALAIYSDPCLCIFT..ASTAYTIVSSIISFYI
O75963        S.VEFDEFKWM.....CVAAWH..REPGYTAFWQLWCALF
HGPRBMY8      QA........AFDERNAICSMIWGASPSYTILSVVSFIVI


ACM4_CHICK    PVVIMTVLYIHISLA.SRSRVRRHKPESRKERKGKSLSFF
YDBM_CAEEL    PTLLMFFVYFKIYQAFAKHRARQIYRQKVIRKHIESTILH
5H1A_HUMAN    PLLLMLVLYGRIFRA.....ARFRIRKTVKKVEK......
5H1A_MOUSE    PLLLMLVLYGRIFRA.....ARFRIRKTVKKVEK......
5H1A_FUGRU    PLILMLVLYGRIFKA.....ARFRIRKTVKKTEKA.....
5HT_LYMST     PMLVMMIIYIRIWLV.....ARSRIRKDKFQMTKARL...
A1AD_HUMAN    PMAVIVMYCRVYVV......A.........RSTTRSL...
A1AD_MOUSE    PMAVIVMYCRVYVV......A.........RSTTRSL...
Q13675        PLAIILVMYCRVYVV......A.........KRESRGL...
Q13729        PLAIILVMYCRVYVV......A.........KRESRGL...
O60451        PLAIILVMYCRVYVV......A.........KRESRGL...
A1AA_RAT      PLAIILVMYCRVYVV......A.........KRESRGL...
O54913        PLTIILVMYCRVYVV......A.........KRESRGL...
A1AA_BOVIN    PLTIILVMYCRVYVV......A.........KRESRGL...
A1AA_CANFA    PLTIILVMYCRVYVV......A.........KRESRGL...
A1AA_RABIT    PLTIILAMYCRVYVV......A.........KRESRGL...
A1AA_HUMAN    PLAIILVMYCRVYVV......A.........KRESRGL...
A1AA_ORYLA    PLAIILAMYCRVYVV......A.........QKESRGL...
O96716        PLLIMLVFYGILFKA......A.........RDQARKI...
O75963        PFLVMLVCYGFIFRV.....ARV.........KARKV...
HGPRBMY8      PLIVMIACYSVVFCAARRQHA.LLYNVKRHSLEVRVKDCV
```

FIG. 6E

ACM4_CHICK	KAPPVKQNNNNSPKRAVEVKEEVRNGKVDDQPSAQTEATG
YDBM_CAEEL	EISHVLPTSDEFAKEEEEEEDSESSGQVENGLGNGNDAI.
5H1A_HUMAN	.........TGADTRHGASPAPQPK..KSVNG....E..
5H1A_MOUSE	.........KGAGTSFGTSSAPPPK..KSLNG....Q..
5H1A_FUGRU	.........KASDMCLTLSPAVFHK..RA.NG....D..
5HT_LYMST	.........KTEETTLVASPKTEYSVVSDCNGCNSPD..
A1AD_HUMAN	.........EA...................GVKRER..
A1AD_MOUSE	.........EA...................GIKREP..
Q13675	.........KS..................GLKTDK..
Q13729	.........KS..................GLKTDK..
O60451	.........KS..................GLKTDK..
A1AA_RAT	.........KS..................GLKTDK..
O54913	.........KS..................GLKTDK..
A1AA_BOVIN	.........KS..................GLKTDK..
A1AA_CANFA	.........KS..................GLKTDK..
A1AA_RABIT	.........KS..................GLKTDK..
A1AA_HUMAN	.........KS..................GLKTDK..
A1AA_ORYLA	.........KE..................GQKIEK..
O96716	.........NA..................LEG....
O75963	.........HCGTVVIVEEDAQRTGRKNSSTSTSSSG..
HGPRBMY8	ENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDG

ACM4_CHICK	QQEEKETSNESSTVSMTQTTKDKPTTEILPAGQGQSPAHP
YDBM_CAEEL	.IEEDECEDEDSDEKRDDHTS...MTTVTATVTGPTEA.P
5H1A_HUMAN	..SGSRNWRLGVESKAGGALCANGAVRQGDDGAAL.EVIE
5H1A_MOUSE	..PGSGDCRRSAENRAVGTPCANGAVRQGEDDATL.EVIE
5H1A_FUGRU	..AVSAEWKRGYKFKP.SSPCANGAVRHGEEMESL.EIIE
5HT_LYMST	..STTEKKKRRAPFKSYG..CSPRPERKKNRAKKLPENAN
A1AD_HUMAN	..GKAS.......................EVVL
A1AD_MOUSE	..GKAS.......................EVVL
Q13675	..SDSE.......................QVTL
Q13729	..SDSE.......................QVTL
O60451	..SDSE.......................QVTL
A1AA_RAT	..SDSE.......................QVTL
O54913	..SDSE.......................QVTL
A1AA_BOVIN	..SDSE.......................QVTL
A1AA_CANFA	..SDSE.......................QVTL
A1AA_RABIT	..SDSE.......................QVTL
A1AA_HUMAN	..SDSE.......................QVIL
A1AA_ORYLA	..SDSE.......................QVIL
O96716	....
O75963	..SRRNAFQGVVYSANQCKALITILVVLGAFMVTWGPYMV
HGPRBMY8	SLKAKEGSTGTSESSVEARGSEEVRESSTVASDGSMEGKE

FIG. 6F

```
ACM4_CHICK   RVNPTSKWSKIKIVTKQTGTESVTAIEIVPAKAGASDHNS
YDBM_CAEEL   YMKREAKISKSVPIEKESAIQKREAKPMRSVMAISYEKVK
5H1A_HUMAN   .VHRVGNSKEHLPL....PSEAGPT....PCAP....ASF
5H1A_MOUSE   .VHRVGNSKGDLPL....PSESGAT....SYVP....ACL
5H1A_FUGRU   .VN..SNSKTHLPL....PN....T....P.QS....SSH
5HT_LYMST    GVNSNSSSSERLKQIQIETAEAFAN....GCAEEASIAML
A1AD_HUMAN   RIHC...............RGAAT....GADGAHGMRSA
A1AD_MOUSE   RIHC...............RGAAT....SAKGNPGTQSS
Q13675       RIHR...............KNAPA......GGSGMASA
Q13729       RIHR...............KNAPA......GGSGMASA
O60451       RIHR...............KNAPA......GGSGMASA
A1AA_RAT     RIHR...............KNVPA......EGGGVSSA
O54913       RIHR...............KNVPA......EGSGVSSA
A1AA_BOVIN   RIHR...............KNAQV......GGSGVTSA
A1AA_CANFA   RIHR...............KNAPV......GGTGVSSA
A1AA_RABIT   RIHR...............KNAPA......GGSGVASA
A1AA_HUMAN   RIHR...............KNAPA......GGSGMASA
A1AA_ORYLA   RMHR...............GNTTV.........SEDEAL
O96716       RLEQ...............EN.........NRGKKISLA
O75963       VI...ASEALWGKSSVSPSLETWAT....WLSFASAVCHP
HGPRBMY8     GSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDINF

ACM4_CHICK   LSNSRPANVARKFASIARSQVRKKRQMAAR..EKKVTRTI
YDBM_CAEEL   RHKNRKERIYRK..SLQR...KPKAISAAK..ERRGVKVL
5H1A_HUMAN   ERK..NERN........AEAKRKMA..LAR..ERKTVKTL
5H1A_MOUSE   ERK..NERT........AEAKRKMA..LAR..ERKTVKTL
5H1A_FUGRU   ENI..NEKT........TGTRRKIA..LAR..ERKTVKTL
5HT_LYMST    ERQCNNGKKISSNDTPYSRTREKIE..LKR..ERKAARTL
A1AD_HUMAN   KG.....HTFRSSLSV.....RLLK..FSR..EKKAAKTL
A1AD_MOUSE   KG.....HTLRSSLSV.....RLLK..FSR..EKKAAKTL
Q13675       KT.....KT...HFSV.....RLLK..FSR..EKKAAKTL
Q13729       KT.....KT...HFSV.....RLLK..FSR..EKKAAKTL
O60451       KT.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_RAT     KN.....KT...HFSV.....RLLK..FSR..EKKAAKTL
O54913       KN.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_BOVIN   KN.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_CANFA   KN.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_RABIT   KN.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_HUMAN   KT.....KT...HFSV.....RLLK..FSR..EKKAAKTL
A1AA_ORYLA   RS.....RT...HFAL.....RLLK..FSR..EKKAAKTL
O96716       K...........................EKKAAKTL
O75963       LIYGLWNKIVRKELLGMCFGDRYYREPFVQ..RQRTSRLF
HGPRBMY8     SEDDVEAVNIPESLPPSRRNSNS.NPPLPRCYQCKAAKVI
```

FIG. 6G

```
ACM4_CHICK   FAILLAFILIWTPYNVMVLINTFC..ETCVPET.VWSIGY
YDBM_CAEEL   GIILGCFTVCWAPFFTMYVLVQFC..KDCSPNAHIEMFIT
5H1A_HUMAN   GIIMGTFILCWLPFFIVALVLPF.CESSSCHMPTLLGAIIN
5H1A_MOUSE   GIIMGTFILCWLPFFIVALVLPF.CESSCHMPELLGAIIN
5H1A_FUGRU   GIIMGTFIFCWLPFFIVALVLPF.CAENCYMPEWLGAVIN
5HT_LYMST    AILTGAFLICWLPFFIIALIGPF.VDPE.GIPPFARSFVL
A1AD_HUMAN   AIVVGVFVLCWFPFFFVLPLGSL.F.PQLKPSEGVFKVIF
A1AD_MOUSE   AIVVGVFVLCWFPFFFVLPLGSL.F.PQLKPSEGVFKVIF
Q13675       GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPSETVFKIVF
Q13729       GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPSETVFKIVF
O60451       GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPSETVFKIVF
A1AA_RAT     GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPSETVFKIVF
O54913       GIVVGCFVLCWLPFFLVMPIGSF.F.PNFKPPETVFKIVF
A1AA_BOVIN   GIVVGCFVLCWLPFFLVMPIGSF.F.PDFRPSETVFKIAF
A1AA_CANFA   GIVVGCFVLCWLPFFLVMPIG~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT   GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPPETVFKIVF
A1AA_HUMAN   GIVVGCFVLCWLPFFLVMPIGSF.F.PDFKPSETVFKIVF
A1AA_ORYLA   GIVVGCFVLCWLPFFLVLPIGSI.F.PAYRPSDTVFKITF
O96716       GIIMGVFILCWLPFF.VVNIVNP.F.CDRCVQPAVFIALT
O75963       SISNRITDLGLSPHLTALMAGGQPLGHSSSTGDTGFSCSQ
HGPRBMY8     FIIIFSYVLSLGPYCFLAVL.AVWVDVETQVPQWVITIII
```

```
ACM4_CHICK   WLCYVNSTINPACYALCNATFKKTFKHLLMCQYRNIGTAR
YDBM_CAEEL   WLGYSNSAMNPIIYTVFNRDYQIALKRLFTSEKKPSSTSR
5H1A_HUMAN   WLGYSNSLLNPVIYAYFNKDFQNAFKKIIKCKFCRQ~~~~
5H1A_MOUSE   WLGYSNSLLNPVIYAYFNKDFQNAFKKIIKCKFCR~~~~~
5H1A_FUGRU   WLGYSNSLLNPIIYAYFNKDFQSAFKKILRCKFHRH~~~~
5HT_LYMST    WLGYFNSLLNPIIYTIFSPEFRSAFQKILFGKYRRGHR~~
A1AD_HUMAN   WLGYFNSCVNPLIYPCSSREFKRAFLRLLRC.QCRRRRR.
A1AD_MOUSE   WLGYFNSCVNPLIYPCSSREFKRAFLRLLRC.QCRRRRR.
Q13675       WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLRRKQS
Q13729       WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLRRKQS
O60451       WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLCRKQS
A1AA_RAT     WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLRRRQS
O54913       WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLRRRQS
A1AA_BOVIN   WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLRRKQS
A1AA_CANFA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT   WLGYLNSCINPIIYPCSSQEFKKAFQNVLKI.QCLRRKQS
A1AA_HUMAN   WLGYLNSCINPIIYPCSSQEFKKAFQNVLRI.QCLCRKQS
A1AA_ORYLA   WLGYFNSCINPIIYLCSNQEFKKAFQSLLGV.HCLRMTPR
O96716       WLGWINSCFNPIIY.AFNKEFRKVFVKMICCHKCRGVTVG
O75963       DSGNLRAL~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8     WLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK
```

FIG. 6H

```
ACM4_CHICK      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
YDBM_CAEEL      V~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_HUMAN      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_MOUSE      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_FUGRU      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5HT_LYMST       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AD_HUMAN      .RRPLWRVY..GHHWRASTSGLRQDCAPSSGDAPPGAPLA
A1AD_MOUSE      .R..LW......PSLRPPLASL..DRRPALRLCPQPAHRT
Q13675          SKH...ALG..YT.LHPPSQAVEGQHKDM.VRIPVGSRET
Q13729          SKH...ALG..YT.LHPPSQAVEGQHKDM.VRIPVGSRET
O60451          SKH...ALG..YT.LHPPSQAVEGQHKDM.VRIPVGSRET
A1AA_RAT        SKH...ALG..YT.LHPPSQALEGQHRDM.VRIPVGSGET
O54913          SKH...ALG..YT.LHPPSQAVEEQHRGM.VRIPVGSGET
A1AA_BOVIN      SKH...TLG..YT.LHAPSHVLEGQHKDL.VRIPVGSAET
A1AA_CANFA      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT      SKH...ALG..YT.LHAPSQALEGQHKDM.VRIPVGSGET
A1AA_HUMAN      SKH...ALG..YT.LHPPSQAVEGQHKDM.VRIPVGSRET
A1AA_ORYLA      AHHHHLSVG..QSQTQGHSLTISLDSKGAPCRLSPSSSVA
O96716          PNHADLNYDPVAMRLKKRGENANGTVNGDANGKANGNIEA
O75963          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8        EDSHPDLPGTEGGTEGKIVPSYDSATFP~~~~~~~~~~~~

ACM4_CHICK      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
YDBM_CAEEL      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_HUMAN      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_MOUSE      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_FUGRU      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5HT_LYMST       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AD_HUMAN      LTALPDPD..PEP...PGTPEMQAPVASRRKPPSA...FR
A1AD_MOUSE      PRGSPSPH..CTPR..PGLRRHAGGAGFGLRPSKASLRLR
Q13675          FYRISKTDGVCEWKFFSSMPRGSARITVSKDQSSCTTART
Q13729          FYRISKTDGVCEWKFFSSMPRGSARITVSKDQSSCTTARG
O60451          FYRISKTDGVCEWKFFSSMPRGSARITVSKDQSSCTTARR
A1AA_RAT        FYKISKTDGVCEWKFFSSMPQGSARITVPKDQSACTTARV
O54913          FYKISKTDGVCEWKFFSSMPQGSARITMPKDQSACTTARV
A1AA_BOVIN      FYKISKTDGVCEWKIFSSLPRGSARMAVARDPSACTTARV
A1AA_CANFA      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT      FYKISKTDGVCEWKFFSSMPRGSARITVPKDQSACTTARV
A1AA_HUMAN      FYRISKTDGVCEWKFFSSMPRGSARITVSKDQSSCTTARV
A1AA_ORYLA      LSRTPSSRDSREWRVFSGGPINSG..PGPTEAGRAKVAKL
O96716          GEGTSSS~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O75963          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

FIG. 6I

```
ACM4_CHICK   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
YDBM_CAEEL   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_HUMAN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_MOUSE   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_FUGRU   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5HT_LYMST    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AD_HUMAN   EWRLLGPFRRPTTQLRAKVSSLSHKIRAGGAQRAEAACAQ
A1AD_MOUSE   EWRLLGPLQRPTTQLRAKVSSLSHKFRSGGARRAETACAL
Q13675       KSRSVTRLECS....GMILAHCN..LRLPGSRDSPASASQ
Q13729       HT.PMT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O60451       GMDCRYFTKNC....REHIKHVN..FMMPPWRKGLEC~~~
A1AA_RAT     RSKSFLQVCCCV.GSSAPRPEEN..HQVPTIKIHTISLGE
O54913       RSKSFLQVCCCV.GSSTPRPEEN..HQVPTIKIHTISLGE
A1AA_BOVIN   RSKSFLQVCCCL.GPSTPSHGEN..HQIPTIKIHTISLSE
A1AA_CANFA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT   RSKSFLQVCCCV.GPSTPNPGEN..HQVPTIKIHTISLSE
A1AA_HUMAN   RSKSFLQVCCCV.GPSTPSLDKN..HQVPTIKVHTISLSE
A1AA_ORYLA   CNKSLHRTCCCILRARTPTQDPAPLGDLPTIKIHQLSLSE
O96716       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O75963       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

ACM4_CHICK   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
YDBM_CAEEL   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_HUMAN   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_MOUSE   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5H1A_FUGRU   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
5HT_LYMST    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AD_HUMAN   RSEVEAVSLGVPHEVAEGATCQAYELADYSNLRETDI~~~
A1AD_MOUSE   RSEVEAVSLNVPQDGAEAVICQAYEPGDLSNLRETDI~~~
Q13675       AAGTTGDVPPGRRHQAQLIFVFLVETGFHHVGQDDLDLLT
Q13729       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O60451       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RAT     NGEEV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O54913       NGEEV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_BOVIN   NGEEV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_CANFA   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_RABIT   NGEEV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_HUMAN   NGEEV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
A1AA_ORYLA   KGESV~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O96716       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
O75963       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HGPRBMY8     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

FIG. 6J

```
ACM4_CHICK        ~
YDBM_CAEEL        ~
5H1A_HUMAN        ~
5H1A_MOUSE        ~
5H1A_FUGRU        ~
5HT_LYMST         ~
A1AD_HUMAN        ~
A1AD_MOUSE        ~
Q13675            S
Q13729            ~
O60451            ~
A1AA_RAT          ~
O54913            ~
A1AA_BOVIN        ~
A1AA_CANFA        ~
A1AA_RABIT        ~
A1AA_HUMAN        ~
A1AA_ORYLA        ~
O96716            ~
O75963            ~
HGPRBMY8          ~
```

D0047 PCT

FIG. 7
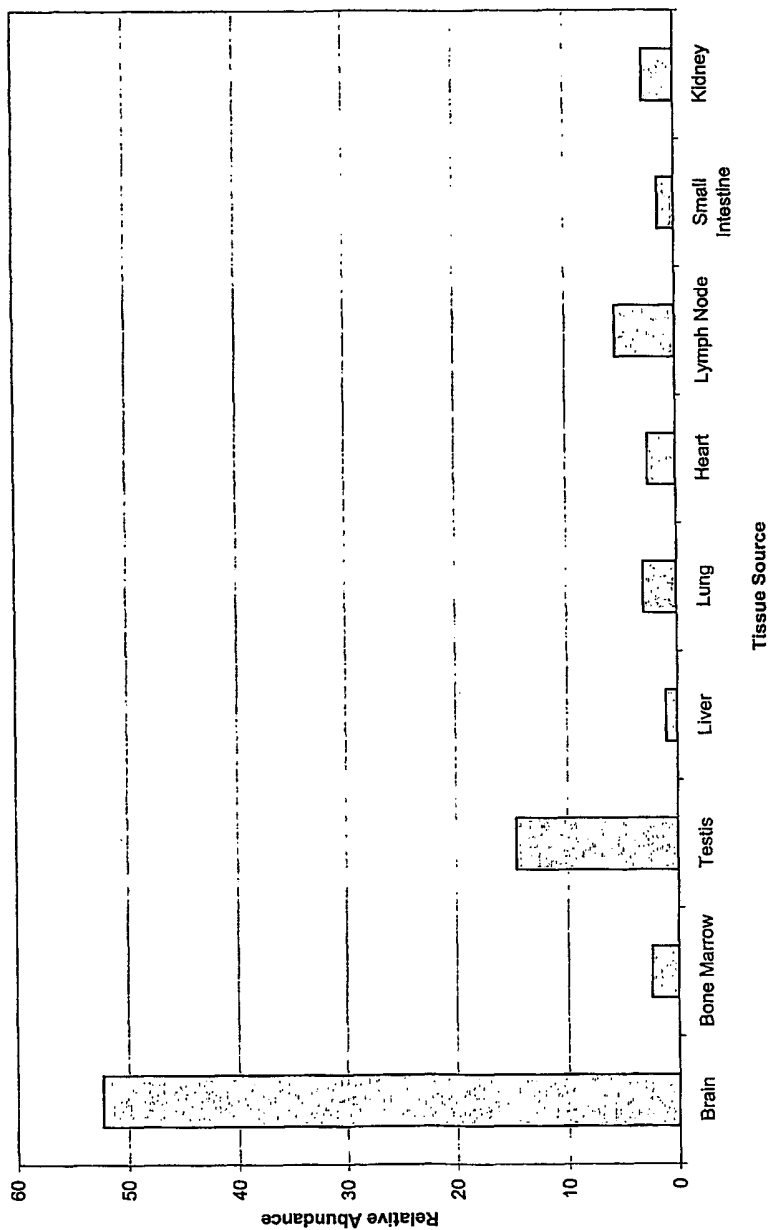
D0047 PCT

FIG. 8
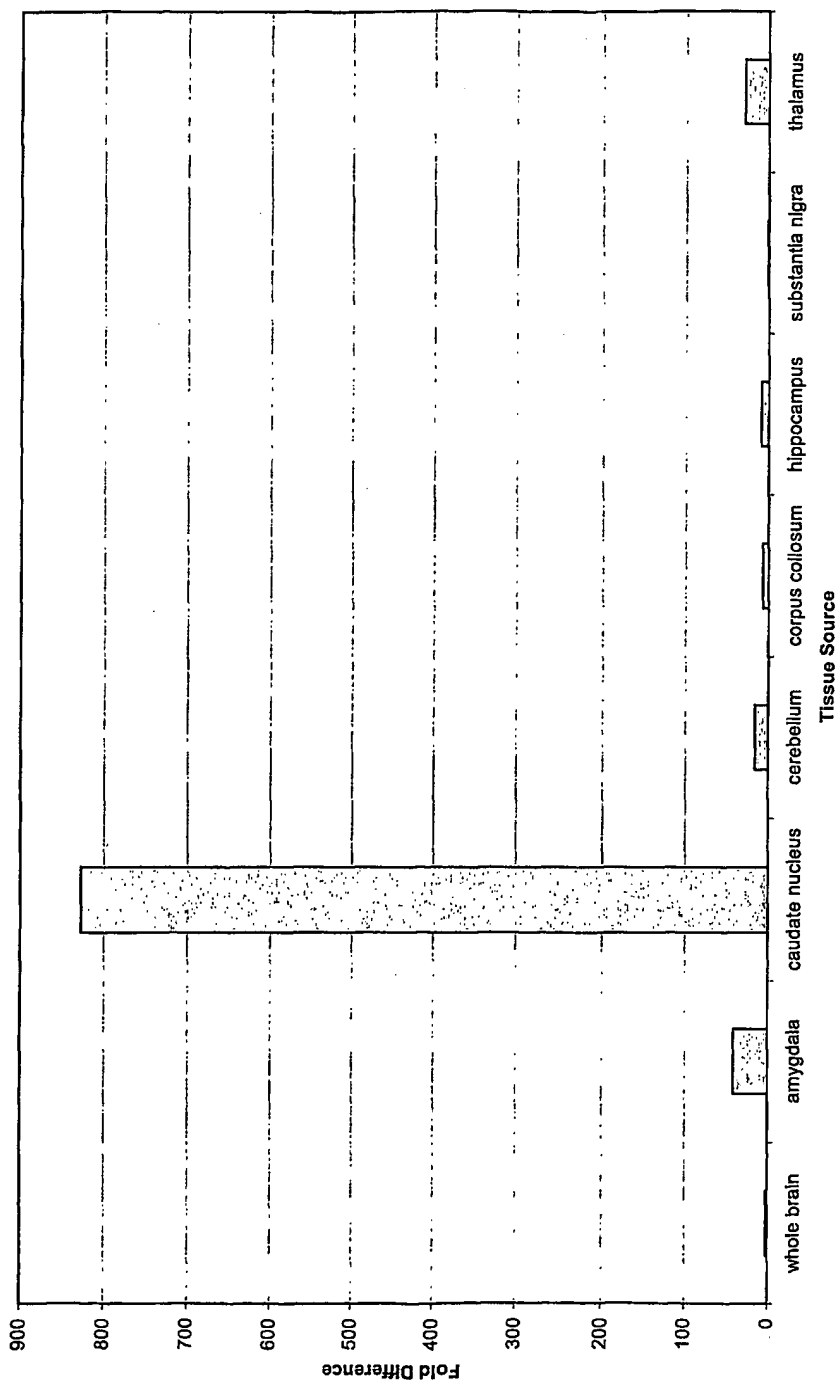
FIG. 9

```
HGPRBMY8   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP
AL390879   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP
AX148250   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP
AX080495   MTSTCTNSTRESNSSHTCMPLSKMPISLAHGIIRSTVLVIFLAASFVGNIVLALVLQRKP

HGPRBMY8   QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN
AL390879   QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN
AX148250   QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN
AX080495   QLLQVTNRFIFNLLVTDLLQISLVAPWVVATSVPLFWPLNSHFCTALVSLTHLFAFASVN

HGPRBMY8   TIVLVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA
AL390879   TIVVVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA
AX148250   TIVVVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA
AX080495   TIVLVSVDRYLSIIHPLSYPSKMTQRRGYLLLYGTWIVAILQSTPPLYGWGQAAFDERNA

HGPRBMY8   LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD
AL390879   LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD
AX148250   LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD
AX080495   LCSMIWGASPSYTILSVVSFIVIPLIVMIACYSVVFCAARRQHALLYNVKRHSLEVRVKD

HGPRBMY8   CVENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEA
AL390879   CVENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEA
AX148250   CVENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEA
AX080495   CVENEDEEGAEKKEEFQDESEFRRQHEGEVKAKEGRMEAKDGSLKAKEGSTGTSESSVEA

HGPRBMY8   RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDI
AL390879   RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDI
AX148250   RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDDMEFGEDDI
AX080495   RGSEEVRESSTVASDGSMEGKEGSTKVEENSMKADKGRTEVNQCSIDLGEDGMEFGEDDI

HGPRBMY8   NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYCFLAV
AL390879   NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYCFLAV
AX148250   NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAKKVIFIIIFSYVLSLGPYCFLAV
AX080495   NFSEDDVEAVNIPESLPPSRRNSNSNPPLPRCYQCKAAKVIFIIIFSYVLSLGPYCFLAV

HGPRBMY8   LAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK
AL390879   LAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK
AX148250   LAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK
AX080495   LAVWVDVETQVPQWVITIIIWLFFLQCCIHPYVYGYMHKTIKKEIQDMLKKFFCKEKPPK

HGPRBMY8   EDSHPDLPGTEGGTEGKIVPSYDSATFP~
AL390879   EDSHPDLPGTEGGTEGKIVPSYDSATFP*
AX148250   EDSHPDLPGTEGGTEGKIVPSYDSATFP*
AX080495   EDSHPDLPGTEGGTEGKIVPSYDSATFP*
```

FIG. 10A

```
AX080495   GCCTGCAACCTGTCYCACGCCCTCTGGCTGTTGCCATGACGTCCACCTGC
HGPRBMY8   ~~~~GCAACCTGTCTCACGCCCTCTGGCTGTTGCCATGACGTCCACCTGC
AL390879   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGACGTCCACCTGC
AX148250   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ATGACGTCCACCTGC

AX080495   ACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCCCTCTC
HGPRBMY8   ACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCCCTCTC
AL390879   ACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCCCTCTC
AX148250   ACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCCCTCTC

AX080495   CAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGG
HGPRBMY8   CAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGG
AL390879   CAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGG
AX148250   CAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGG

AX080495   TTATCTTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTG
HGPRBMY8   TTATCTTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTG
AL390879   TTATCTTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTG
AX148250   TTATCTTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTG

AX080495   TTGCAGCGCAAGCCGCAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAA
HGPRBMY8   TTGCAGCGCAAGCCGCAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAA
AL390879   TTGCAGCGCAAGCCGCAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAA
AX148250   TTGCAGCGCAAGCCGCAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAA

AX080495   CCTCCTCGTCACCGACCTGCTGCAGATTTCGCTCGTGGCCCCCTGGGTGG
HGPRBMY8   CCTCCTCGTCACCGACCTGCTGCAGATTTCGCTCGTGGCCCCCTGGGTGG
AL390879   CCTCCTCGTCACCGACCTGCTGCAGATTTCGCTCGTGGCCCCCTGGGTGG
AX148250   CCTCCTCGTCACCGACCTGCTGCAGATTTCGCTCGTGGCCCCCTGGGTGG

AX080495   TGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAACAGCCACTTCTGCACG
HGPRBMY8   TGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAACAGCCACTTCTGCACG
AL390879   TGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAACAGCCACTTCTGCACG
AX148250   TGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAACAGCCACTTCTGCACG

AX080495   GCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAACACCAT
HGPRBMY8   GCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAACACCAT
AL390879   GCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAACACCAT
AX148250   GCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAACACCAT

AX080495   TGTCTTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCT
HGPRBMY8   TGTCTTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCT
AL390879   TGTCGTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCT
AX148250   TGTCGTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCT

AX080495   ACCCGTCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACC
HGPRBMY8   ACCCGTCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACC
AL390879   ACCCGTCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACC
AX148250   ACCCGTCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACC
```

FIG. 10B

```
AX080495   TGGATTGTGGCCATCCTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCA
HGPRBMY8   TGGATTGTGGCCATCCTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCA
AL390879   TGGATTGTGGCCATCCTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCA
AX148250   TGGATTGTGGCCATCCTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCA

AX080495   GGCTGCCTTTGATGAGCGCAATGCTCTCTGCTCCATGATCTGGGGGGCCA
HGPRBMY8   GGCTGCCTTTGATGAGCGCAATGCTCTCTGCTCCATGATCTGGGGGGCCA
AL390879   GGCTGCCTTTGATGAGCGCAATGCTCTCTGCTCCATGATCTGGGGGGCCA
AX148250   GGCTGCCTTTGATGAGCGCAATGCTCTCTGCTCCATGATCTGGGGGGCCA

AX080495   GCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTCATCGTCATTCCACTG
HGPRBMY8   GCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTCATCGTCATTCCACTG
AL390879   GCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTCATCGTCATTCCACTC
AX148250   GCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTCATCGTCATTCCACTG

AX080495   ATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGGAGGCA
HGPRBMY8   ATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGGAGGCA
AL390879   ATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGGAGGCA
AX148250   ATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGGAGGCA

AX080495   GCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCA
HGPRBMY8   GCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCA
AL390879   GCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCA
AX148250   GCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCA

AX080495   AGGACTGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAG
HGPRBMY8   AGGACTGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAG
AL390879   AGGACTGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAG
AX148250   AGGACTGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAG

AX080495   TTCCAGGATGAGAGTGAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGC
HGPRBMY8   TTCCAGGATGAGAGTGAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGC
AL390879   TTCCAGGATGAGAGTGAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGC
AX148250   TTCCAGGATGAGAGTGAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGC

AX080495   CAAGGAGGGCAGAATGGAAGCCAAGGACGGCAGCCTGAAGGCCAAGGAAG
HGPRBMY8   CAAGGAGGGCAGAATGGAAGCCAAGGACGGCAGCCTGAAGGCCAAGGAAG
AL390879   CAAGGAGGGCAGAATGGAAGCCAAGGACGGCAGCCTGAAGGCCAAGGAAG
AX148250   CAAGGAGGGCAGAATGGAAGCCAAGGACGGCAGCCTGAAGGCCAAGGAAG

AX080495   GAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCCAGGGGCAGCGAGGAG
HGPRBMY8   GAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCCAGGGGCAGCGAGGAG
AL390879   GAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCCAGGGGCAGCGAGGAG
AX148250   GAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCCAGGGGCAGCGAGGAG

AX080495   GTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGTAAGGA
HGPRBMY8   GTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGTAAGGA
AL390879   GTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGTAAGGA
AX148250   GTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGTAAGGA
```

FIG. 10C

```
AX080495    AGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCA
HGPRBMY8    AGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCA
AL390879    AGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCA
AX148250    AGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCA

AX080495    CAGAGGTCAACCAGTGCAGCATTGACTTGGGTGAAGATGGCATGGAGTTT
HGPRBMY8    CAGAGGTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTT
AL390879    CAGAGGTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTT
AX148250    CAGAGGTCAACCAGTGCAGCATTGACTTGGGTGAAGATGACATGGAGTTT

AX080495    GGTGAAGACGACATCAATTTCAGTGAGGATGACGTCGAGGCAGTGAACAT
HGPRBMY8    GGTGAAGACGACATCAATTTCAGTGAGGATGACGTCGAGGCAGTGAACAT
AL390879    GGTGAAGACGACATCAATTTCAGTGAGGATGACGTCGAGGCAGTGAACAT
AX148250    GGTGAAGACGACATCAATTTCAGTGAGGATGACGTCGAGGCAGTGAACAT

AX080495    CCCGGAGAGCCTCCCACCCAGTCGTCGTAACAGCAACAGCAACCCTCCTC
HGPRBMY8    CCCGGAGAGCCTCCCACCCAGTCGTCGTAACAGCAACAGCAACCCTCCTC
AL390879    CCCGGAGAGCCTCCCACCCAGTCGTCGTAACAGCAACAGCAACCCTCCTC
AX148250    CCCGGAGAGCCTCCCACCCAGTCGTCGTAACAGCAACAGCAACCCTCCTC

AX080495    TGCCCAGGTGCTACCAGTGCAAAGCTGCTAAAGTGATCTTCATCATCATT
HGPRBMY8    TGCCCAGGTGCTACCAGTGCAAAGCTGCTAAAGTGATCTTCATCATCATT
AL390879    TGCCCAGGTGCTACCAGTGCAAAGCTGCTAAAGTGATCTTCATCATCATT
AX148250    TGCCCAGGTGCTACCAGTGCAAAGCTAAGAAAGTGATCTTCATCATCATT

AX080495    TTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTCCTGGC
HGPRBMY8    TTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTCCTGGC
AL390879    TTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTCCTGGC
AX148250    TTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTCCTGGC

AX080495    CGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAA
HGPRBMY8    CGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAA
AL390879    CGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAA
AX148250    CGTGTGGGTGGATGTCGAAACCCAGGTACCCCAGTGGGTGATCACCATAA

AX080495    TCATCTGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGC
HGPRBMY8    TCATCTGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGC
AL390879    TCATCTGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGC
AX148250    TCATCTGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGC

AX080495    TACATGCACAAGACCATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTT
HGPRBMY8    TACATGCACAAGACCATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTT
AL390879    TACATGCACAAGACCATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTT
AX148250    TACATGCACAAGACCATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTT

AX080495    CTTCTGCAAGGAAAAGCCCCCGAAAGAAGATAGCCACCCAGACCTGCCCG
HGPRBMY8    CTTCTGCAAGGAAAAGCCCCCGAAAGAAGATAGCCACCCAGACCTGCCCG
AL390879    CTTCTGCAAGGAAAAGCCCCCGAAAGAAGATAGCCACCCAGACCTGCCCG
AX148250    CTTCTGCAAGGAAAAGCCCCCGAAAGAAGATAGCCACCCAGACCTGCCCG
```

FIG. 10D

```
AX080495  GAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCTTCCTACGATTCTGCT
HGPRBMY8  GAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCTTCCTACGATTCTGCT
AL390879  GAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCTTCCTACGATTCTGCT
AX148250  GAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCTTCCTACGATTCTGCT

AX080495  ACTTTTCCTTGAAGTTAGTTCTAAGGCAAACCTTGAAAATCAGTCCTTCA
HGPRBMY8  ACTTTTCCTTGAAGTTAGTTCTAAGGCAAACCTT~~~~~~~~~~~~~~~~~
AL390879  ACTTTTCCTTGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
AX148250  ACTTTTCCTTGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

AX080495  GCCACAGCTATTTAGAGCTTTAAAACTACCAGGTTCAATCACTGGTTATG
HGPRBMY8  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
AL390879  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
AX148250  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

AX080495  CTTTCTGTG
HGPRBMY8  ~~~~~~~~~
AL390879  ~~~~~~~~~
AX148250  ~~~~~~~~~
```

FIG. 11
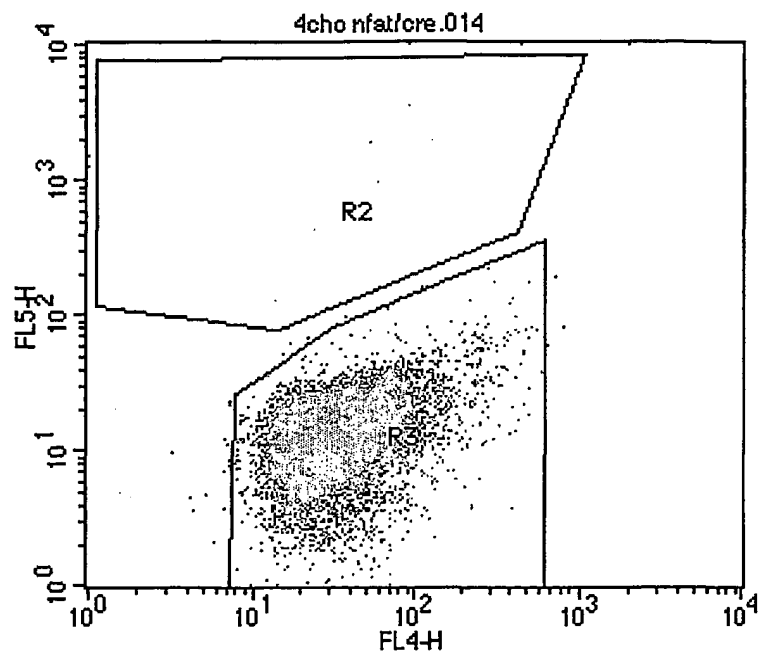
FIG. 12
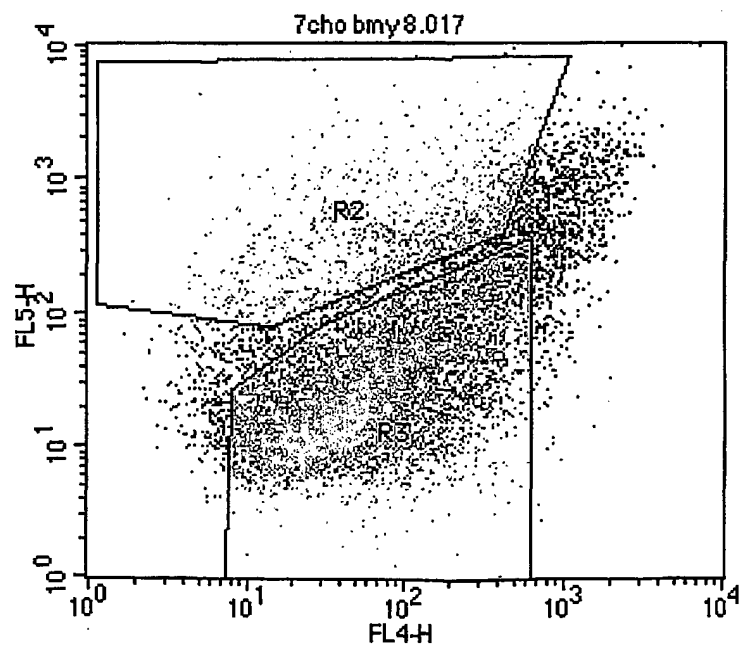
FIG. 13
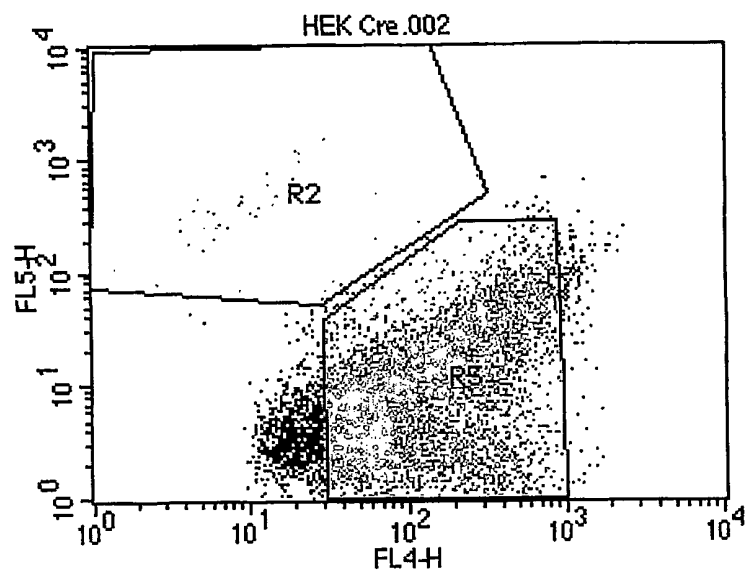
FIG. 14
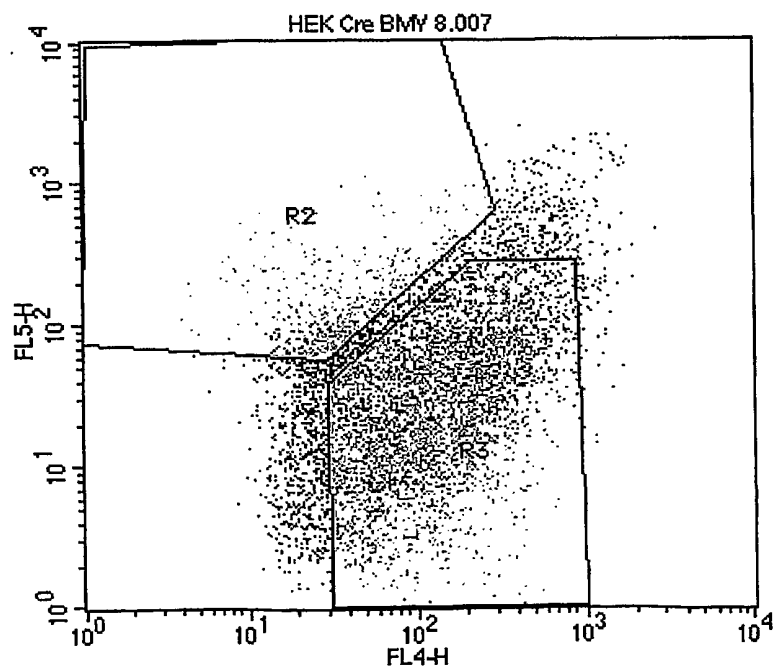
FIG. 15

a. CHO-NFAT G alpha 15 (Fluorescent vs. Bright Field)
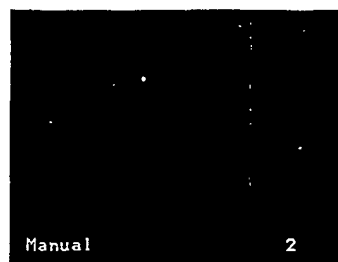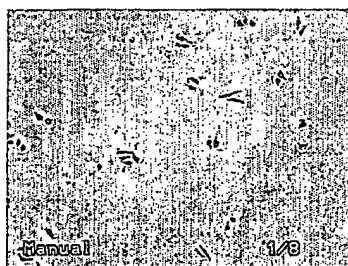
b. CHO-NFAT/ G alpha 15 HGPRBMY8 (Fluorescent vs. Bright Field)
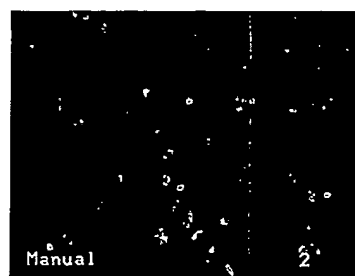
FIG. 16

a. CHO-NFAT/CRE
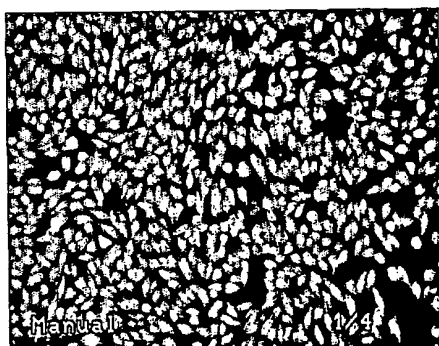
b. CHO-NFAT/CRE + F/T/P
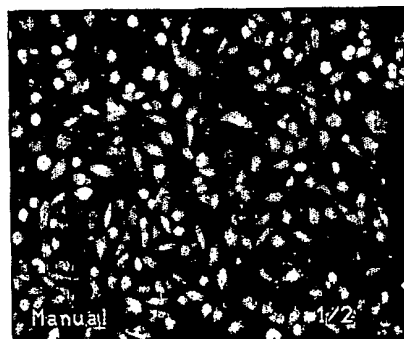
c. CHO-NFAT/CRE oGPCR-Intermediate
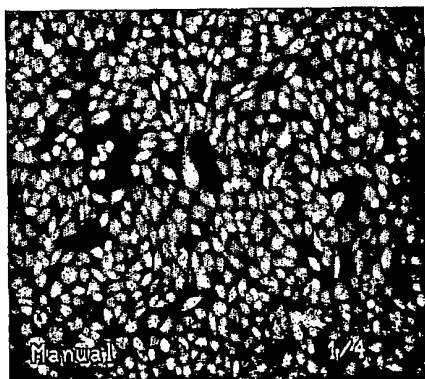
d. CHO-NFAT/CRE oGPCR high

FIG. 17

Normalized GPCR Levels

Tissue Types
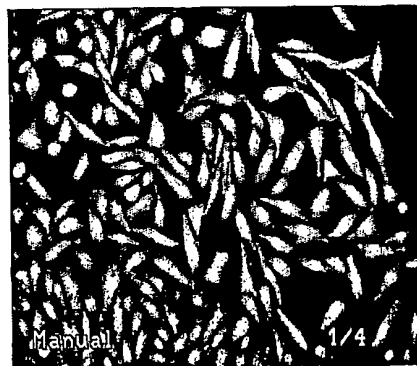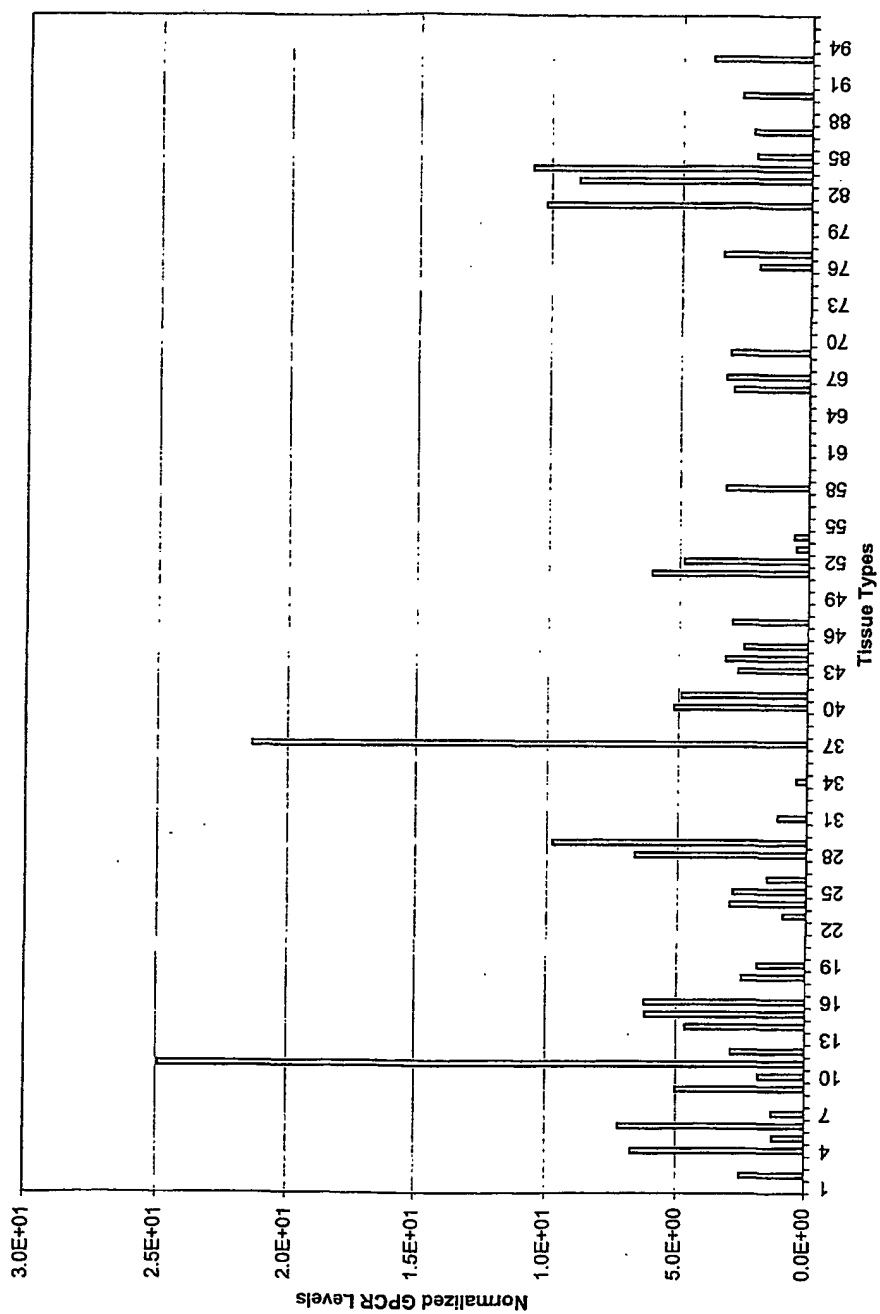
FIG. 18A

```
  1  ATGACGTCCACCTGCACCAACAGCACGCGCGAGAGTAACAGCAGCCACACGTGCATGCCC   60
  1   M  T  S  C  T  N  S  T  R  E  S  N  S  S  H  T  C  M  P      20


 61  CTCTCCAAAATGCCCATCAGCCTGGCCCACGGCATCATCCGCTCAACCGTGCTGGTTATC  120
 21   L  S  K  M  P  I  S  L  A  H  G  I  I  R  S  T  V  L  V  I   40


121  TTCCTCGCCGCCTCTTTCGTCGGCAACATAGTGCTGGCGCTAGTGTTGCAGCGCAAGCCG  180
 41   F  L  A  A  S  F  V  G  N  I  V  L  A  L  V  L  Q  R  K  P   60


181  CAGCTGCTGCAGGTGACCAACCGTTTTATCTTTAACCTCCTCGTCACCGACCTGCTGCAG  240
 61   Q  L  L  Q  V  T  N  R  F  I  F  N  L  L  V  T  D  L  L  Q   80


241  ATTTCGCTCGTGGCCCCCTGGGTGGTGGCCACCTCTGTGCCTCTCTTCTGGCCCCTCAAC  300
 81   I  S  L  V  A  P  W  V  V  A  T  S  V  P  L  F  W  P  L  N  100


301  AGCCACTTCTGCACGGCCCTGGTTAGCCTCACCCACCTGTTCGCCTTCGCCAGCGTCAAC  360
101   S  H  F  C  T  A  L  V  S  L  T  H  L  F  A  F  A  S  V  N  120


361  ACCATTGTCNTGGTGTCAGTGGATCGCTACTTGTCCATCATCCACCCTCTCTCCTACCCG  420
121   T  I  V  X  V  S  V  D  R  Y  L  S  I  I  H  P  L  S  Y  P  140
```
(underline under X at position 124)

```
421  TCCAAGATGACCCAGCGCCGCGGTTACCTGCTCCTCTATGGCACCTGGATTGTGGCCATC  480
141   S  K  M  T  Q  R  R  G  Y  L  L  L  Y  G  T  W  I  V  A  I  160


481  CTGCAGAGCACTCCTCCACTCTACGGCTGGGGCCAGGCTGCCTTTGATGAGCGCAATGCT  540
161   L  Q  S  T  P  P  L  Y  G  W  G  Q  A  A  F  D  E  R  N  A  180


541  CTCTGCTCCATGATCTGGGGGGCCAGCCCCAGCTACACTATTCTCAGCGTGGTGTCCTTC  600
181   L  C  S  M  I  W  G  A  S  P  S  Y  T  I  L  S  V  V  S  F  200


601  ATCGTCATTCCACTGATTGTCATGATTGCCTGCTACTCCGTGGTGTTCTGTGCAGCCCGG  660
201   I  V  I  P  L  I  V  M  I  A  C  Y  S  V  V  F  C  A  A  R  220


661  AGGCAGCATGCTCTGCTGTACAATGTCAAGAGACACAGCTTGGAAGTGCGAGTCAAGGAC  720
221   R  Q  H  A  L  L  Y  N  V  K  R  H  S  L  E  V  R  V  K  D  240


721  TGTGTGGAGAATGAGGATGAAGAGGGAGCAGAGAAGAAGGAGGAGTTCCAGGATGAGAGT  780
241   C  V  E  N  E  D  E  E  G  A  E  K  K  E  E  F  Q  D  E  S  260


781  GAGTTTCGCCGCCAGCATGAAGGTGAGGTCAAGGCCAAGGAGGGCAGAATGGAAGCCAAG  840
261   E  F  R  R  Q  H  E  G  E  V  K  A  K  E  G  R  M  E  A  K  280


841  GACGGCAGCCTGAAGGCCAAGGAAGGAAGCACGGGGACCAGTGAGAGTAGTGTAGAGGCC  900
281   D  G  S  L  K  A  K  E  G  S  T  G  T  S  E  S  S  V  E  A  300
```

FIG. 18B

```
 901  AGGGGCAGCGAGGAGGTCAGAGAGAGCAGCACGGTGGCCAGCGACGGCAGCATGGAGGGT   960
 301   R   G   S   E   E   V   R   E   S   S   T   V   A   S   D   G   S   M   E   G    320

 961  AAGGAAGGCAGCACCAAAGTTGAGGAGAACAGCATGAAGGCAGACAAGGGTCGCACAGAG  1020
 321   K   E   G   S   T   K   V   E   E   N   S   M   K   A   D   K   G   R   T   E    340

1021  GTCAACCAGTGCAGCATTGACTTGGGTGAAGATGNCATGGAGTTTGGTGAAGACGACATC  1080
 341   V   N   Q   C   S   I   D   L   G   E   D   X̲   M   E   F   G   E   D   D   I    360

1081  AATTTCAGTGAGGATGACGTCGAGGCAGTGAACATCCCGGAGAGCCTCCCACCCAGTCGT  1140
 361   N   F   S   E   D   D   V   E   A   V   N   I   P   E   S   L   P   P   S   R    380

1141  CGTAACAGCAACAGCAACCCTCCTCTGCCCAGGTGCTACCAGTGCAAAGCTNNNAAAGTG  1200
 381   R   N   S   N   S   N   P   P   L   P   R   C   Y   Q   C   K   A   X̲   K   V    400

1201  ATCTTCATCATCATTTTCTCCTATGTGCTATCCCTGGGGCCCTACTGCTTTTTAGCAGTC  1260
 401   I   F   I   I   I   F   S   Y   V   L   S   L   G   P   Y   C   F   L   A   V    420

1261  CTGGCCGTGTGGGTGGATGTCGAAACCCAGGTACCCAGTGGGTGATCACCATAATCATC  1320
 421   L   A   V   W   V   D   V   E   T   Q   V   P   Q   W   V   I   T   I   I   I    440

1321  TGGCTTTTCTTCCTGCAGTGCTGCATCCACCCCTATGTCTATGGCTACATGCACAAGACC  1380
 441   W   L   F   F   L   Q   C   C   I   H   P   Y   V   Y   G   Y   M   H   K   T    460

1381  ATTAAGAAGGAAATCCAGGACATGCTGAAGAAGTTCTTCTGCAAGGAAAAGCCCCCGAAA  1440
 461   I   K   K   E   I   Q   D   M   L   K   K   F   F   C   K   E   K   P   P   K    480

1441  GAAGATAGCCACCCAGACCTGCCCGGAACAGAGGGTGGGACTGAAGGCAAGATTGTCCCT  1500
 481   E   D   S   H   P   D   L   P   G   T   E   G   G   T   E   G   K   I   V   P    500

1501  TCCTACGATTCTGCTACTTTTCCTTGA  1527
 501   S   Y   D   S   A   T   F   P    508
```

SEQUENCE LISTING

<110> BATTAGLINO, PETER
      FEDER, JOHN N
      MINTIER, GABE
      NELSON, THOMAS C
      RAMANATHAN, CHANDRA S
      WESTPHAL, RYAN
      CACACE, ANGELA
      BARBER, LAUREN
      HAWKEN, DONALD R
      KORNACKER, MICHAEL G

<120> A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY8,
      EXPRESSED HIGHLY IN BRAIN

<130> D0047NP

<140> TBA
<141> 2001-11-13

<150> 60/317166
<151> 2001-09-04

<150> 60/308285
<151> 2001-07-27

<150> 60/268581
<151> 2001-02-14

<150> 60/248285
<151> 2000-11-14

<160> 102

<170> PatentIn Ver. 2.1

<210> 1
<211> 1527
<212> DNA
<213> Homo sapiens

<400> 1
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac gtgcatgccc 60
ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca gcgcaagccg 180
cagctgctgc aggtgaccaa ccgtttttatc tttaacctcc tcgtcaccga cctgctgcag 240
atttcgctcg tggcccctg ggtggtggcc acctctgtgc ctctcttctg gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc cagcgtcaac 360
accattgtct tggtgtcagt ggatcgctac ttgtccatca tccaccctct ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg cacctggat tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt ggtgtccttc 600
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct ggaagtgcg agtcaaggac 720
tgtgtggaga tgaggatga agagggagca gagaagaagg aggagttcca ggatgagagt 780
gagtttcgcc gccagcatga aggtgaggtc aaggccaagg agggcagaat ggaagccaag 840
gacggcagcc tgaaggccaa ggaaggaagc acggggacca gtgagagtag tgtagaggcc 900
aggggcagcg aggaggtcag agagagcagc acggtggcca gcgacggcag catggagggt 960

aaggaaggca gcaccaaagt tgaggagaac agcatgaagg cagacaaggg tcgcacagag
1020
gtcaaccagt gcagcattga cttgggtgaa gatgacatgg agtttggtga agacgacatc
1080
aatttcagtg aggatgacgt cgaggcagtg aacatcccgg agagcctccc acccagtcgt
1140
cgtaacagca acagcaaccc tcctctgccc aggtgctacc agtgcaaagc tgctaaagtg
1200
atcttcatca tcattttctc ctatgtgcta tccctggggc cctactgctt tttagcagtc
1260
ctggccgtgt gggtggatgt cgaaacccag gtaccccagt gggtgatcac cataatcatc
1320
tggctttttct tcctgcagtg ctgcatccac ccctatgtct atggctacat gcacaagacc
1380
attaagaagg aaatccagga catgctgaag aagttcttct gcaaggaaaa gcccccgaaa
1440
gaagatagcc acccagacct gcccggaaca gagggtggga ctgaaggcaa gattgtccct
1500
tcctacgatt ctgctacttt tccttga
1527

<210> 2
<211> 508
<212> PRT
<213> Homo sapiens

<400> 2
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
 1   5   10   15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
            20   25   30

Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
         35   40   45

Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
      50   55   60

Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
 65   70   75   80

Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
               85   90   95

Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
            100   105   110

Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Leu Val Ser Val Asp
            115   120   125

Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
         130   135   140

Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145   150   155   160

Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp

```
                  165                    170                    175
```

Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
            180   185   190

Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
        195   200   205

Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
    210   215   220

Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225   230   235   240

Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
                245   250   255

Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
                260   265   270

Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
            275   280   285

Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
290   295   300

Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305   310   315   320

Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys
                325   330   335

Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Asp
            340   345   350

Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
        355   360   365

Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
    370   375   380

Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Ala Lys Val
385   390   395   400

Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
                405   410   415

Phe Leu Ala Val Leu Ala Val Trp Val Asp Val Glu Thr Gln Val Pro
            420   425   430

Gln Trp Val Ile Thr Ile Ile Ile Trp Leu Phe Phe Leu Gln Cys Cys
        435   440   445

Ile His Pro Tyr Val Tyr Gly Tyr Met His Lys Thr Ile Lys Lys Glu
    450   455   460

Ile Gln Asp Met Leu Lys Lys Phe Phe Cys Lys Glu Lys Pro Pro Lys
465   470   475   480

3

```
Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly
                485             490             495

Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
            500             505


<210> 3
<211> 31
<212> DNA
<213> Homo sapiens

<400> 3
gcaacctgtc tcacgccctc tggctgttgc c                                31


<210> 4
<211> 22
<212> DNA
<213> Homo sapiens

<400> 4
agttagttct aaggcaaacc tt                                          22


<210> 5
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   HGPRBMY8
      sense primer

<400> 5
ggccgaattc gcaacctgtc tcacgccctc tgg                              33


<210> 6
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   HGPRBMY8
      anti-sense primer

<400> 6
ggccgaattc ggacagttca aggtttgcct tagaac                           36


<210> 7
<211> 490
<212> PRT
<213> Gallus gallus

<400> 7
Met His Asn Leu Ser Ala Gln Pro Trp Gln Ala Lys Met Ala Asn Leu
  1               5               10              15
```

4

```
Thr Tyr Asp Asn Val Thr Leu Ser Asn Arg Ser Glu Val Ala Ile Gln
            20                  25                  30

Pro Pro Thr Asn Tyr Lys Thr Val Glu Leu Val Phe Ile Ala Thr Val
            35                  40                  45

Thr Gly Ser Leu Ser Leu Val Thr Val Val Gly Asn Ile Leu Val Met
    50                  55                  60

Leu Ser Ile Lys Val Asn Arg Gln Leu Gln Thr Val Asn Asn Tyr Phe
65                  70                  75                  80

Leu Phe Ser Leu Ala Cys Ala Asp Leu Ile Ile Gly Val Phe Ser Met
            85                  90                  95

Asn Leu Tyr Thr Val Tyr Ile Ile Lys Gly Tyr Trp Pro Leu Gly Ala
            100                 105                 110

Val Val Cys Asp Leu Trp Leu Ala Leu Asp Tyr Val Val Ser Asn Ala
            115                 120                 125

Ser Val Met Asn Leu Leu Ile Ile Ser Phe Asp Arg Tyr Phe Cys Val
    130                 135                 140

Thr Lys Pro Leu Thr Tyr Pro Ala Arg Arg Thr Thr Lys Met Ala Gly
145                 150                 155                 160

Leu Met Ile Ala Ala Ala Trp Ile Leu Ser Phe Ile Leu Trp Ala Pro
            165                 170                 175

Ala Ile Leu Phe Trp Gln Phe Ile Val Gly Lys Arg Thr Val His Glu
            180                 185                 190

Arg Glu Cys Tyr Ile Gln Phe Leu Ser Asn Pro Ala Val Thr Phe Gly
            195                 200                 205

Thr Ala Ile Ala Ala Phe Tyr Leu Pro Val Val Ile Met Thr Val Leu
    210                 215                 220

Tyr Ile His Ile Ser Leu Ala Ser Arg Ser Arg Val Arg Arg His Lys
225                 230                 235                 240

Pro Glu Ser Arg Lys Glu Arg Lys Gly Lys Ser Leu Ser Phe Phe Lys
            245                 250                 255

Ala Pro Pro Val Lys Gln Asn Asn Asn Ser Pro Lys Arg Ala Val
            260                 265                 270

Glu Val Lys Glu Glu Val Arg Asn Gly Lys Val Asp Asp Gln Pro Ser
            275                 280                 285

Ala Gln Thr Glu Ala Thr Gly Gln Gln Glu Glu Lys Glu Thr Ser Asn
    290                 295                 300

Glu Ser Ser Thr Val Ser Met Thr Gln Thr Thr Lys Asp Lys Pro Thr
305                 310                 315                 320

Thr Glu Ile Leu Pro Ala Gly Gln Gly Gln Ser Pro Ala His Pro Arg
```

```
                 325                      330                      335
    Val Asn Pro Thr Ser Lys Trp Ser Lys Ile Lys Ile Val Thr Lys Gln
                 340                      345                      350

    Thr Gly Thr Glu Ser Val Thr Ala Ile Glu Ile Val Pro Ala Lys Ala
                 355                      360                      365

    Gly Ala Ser Asp His Asn Ser Leu Ser Asn Ser Arg Pro Ala Asn Val
                 370                      375                      380

    Ala Arg Lys Phe Ala Ser Ile Ala Arg Ser Gln Val Arg Lys Lys Arg
    385                      390                      395                      400

    Gln Met Ala Ala Arg Glu Lys Lys Val Thr Arg Thr Ile Phe Ala Ile
                 405                      410                      415

    Leu Leu Ala Phe Ile Leu Thr Trp Thr Pro Tyr Asn Val Met Val Leu
                 420                      425                      430

    Ile Asn Thr Phe Cys Glu Thr Cys Val Pro Glu Thr Val Trp Ser Ile
                 435                      440                      445

    Gly Tyr Trp Leu Cys Tyr Val Asn Ser Thr Ile Asn Pro Ala Cys Tyr
    450                      455                      460

    Ala Leu Cys Asn Ala Thr Phe Lys Lys Thr Phe Lys His Leu Leu Met
    465                      470                      475                      480

    Cys Gln Tyr Arg Asn Ile Gly Thr Ala Arg
                 485                      490
```

```
<210> 8
<211> 488
<212> PRT
<213> Caenorhabditis elegans

<400> 8
Met Cys Phe Ala Glu Lys Gly Glu Gly Ala Gly Glu Asp Val Asp His
1               5                   10                  15

His Ser Leu Phe Cys Pro Lys Lys Leu Val Gly Asn Leu Lys Gly Phe
            20                  25                  30

Ile Arg Asn Gln Tyr His Gln His Glu Thr Ile Gln Ile Leu Lys Gly
        35                  40                  45

Ser Ala Leu Phe Leu Leu Val Leu Trp Thr Ile Phe Ala Asn Ser Leu
    50                  55                  60

Val Phe Ile Val Leu Tyr Lys Asn Pro Arg Leu Gln Thr Val Pro Asn
65                  70                  75                  80

Leu Leu Val Gly Asn Leu Ala Phe Ser Asp Leu Ala Leu Gly Leu Ile
            85                  90                  95

Val Leu Pro Leu Ser Ser Val Tyr Ala Ile Ala Gly Glu Trp Val Phe
            100                 105                 110
```

6

Pro Asp Ala Leu Cys Glu Val Phe Val Ser Ala Asp Ile Leu Cys Ser
      115   120   125

Thr Ala Ser Ile Trp Asn Leu Ser Ile Val Gly Leu Asp Arg Tyr Trp
    130   135   140

Ala Ile Thr Ser Pro Val Ala Tyr Met Ser Lys Arg Asn Lys Arg Thr
145   150   155   160

Ala Gly Ile Met Ile Leu Ser Val Trp Ile Ser Ser Ala Leu Ile Ser
              165   170   175

Leu Ala Pro Leu Leu Gly Trp Lys Gln Thr Ala Gln Thr Pro Asn Leu
          180   185   190

Ile Tyr Glu Lys Asn Asn Thr Val Arg Gln Cys Thr Phe Leu Asp Leu
      195   200   205

Pro Ser Tyr Thr Val Tyr Ser Ala Thr Gly Ser Phe Phe Ile Pro Thr
    210   215   220

Leu Leu Met Phe Phe Val Tyr Phe Lys Ile Tyr Gln Ala Phe Ala Lys
225   230   235   240

His Arg Ala Arg Gln Ile Tyr Arg Gln Lys Val Ile Arg Lys His Ile
              245   250   255

Glu Ser Thr Ile Leu His Glu Ile Ser His Val Leu Pro Thr Ser Asp
          260   265   270

Glu Phe Ala Lys Glu Glu Glu Glu Glu Glu Asp Ser Glu Ser Ser Gly
      275   280   285

Gln Val Glu Asn Gly Leu Gly Asn Gly Asn Asp Ala Ile Ile Glu Glu
    290   295   300

Asp Glu Cys Glu Asp Glu Asp Ser Asp Glu Lys Arg Asp Asp His Thr
305   310   315   320

Ser Met Thr Thr Val Thr Ala Thr Val Thr Gly Pro Thr Glu Ala Pro
              325   330   335

Tyr Met Lys Arg Glu Ala Lys Ile Ser Lys Ser Val Pro Ile Glu Lys
          340   345   350

Glu Ser Ala Ile Gln Lys Arg Glu Ala Lys Pro Met Arg Ser Val Met
      355   360   365

Ala Ile Ser Tyr Glu Lys Val Lys Arg His Lys Asn Arg Lys Glu Arg
    370   375   380

Ile Tyr Arg Lys Ser Leu Gln Arg Lys Pro Lys Ala Ile Ser Ala Ala
385   390   395   400

Lys Glu Arg Arg Gly Val Lys Val Leu Gly Ile Ile Leu Gly Cys Phe
              405   410   415

Thr Val Cys Trp Ala Pro Phe Phe Thr Met Tyr Val Leu Val Gln Phe

7

```
                420                      425                      430

    Cys Lys Asp Cys Ser Pro Asn Ala His Ile Glu Met Phe Ile Thr Trp
            435                 440                 445

    Leu Gly Tyr Ser Asn Ser Ala Met Asn Pro Ile Ile Tyr Thr Val Phe
            450                 455                 460

    Asn Arg Asp Tyr Gln Ile Ala Leu Lys Arg Leu Phe Thr Ser Glu Lys
    465                 470                 475                 480

    Lys Pro Ser Ser Thr Ser Arg Val
                    485


    <210> 9
    <211> 423
    <212> PRT
    <213> Homo sapiens

    <400> 9
    Met Asp Leu Arg Ala Thr Ser Ser Asn Asp Ser Asn Ala Thr Ser Gly
    1               5                   10                  15

    Tyr Ser Asp Thr Ala Ala Val Asp Trp Asp Glu Gly Glu Asn Ala Thr
                20                  25                  30

    Gly Ser Gly Ser Leu Pro Asp Pro Glu Leu Ser Tyr Gln Ile Ile Thr
            35                  40                  45

    Ser Leu Phe Leu Gly Ala Leu Ile Leu Cys Ser Ile Phe Gly Asn Ser
        50                  55                  60

    Cys Val Val Ala Ala Ile Ala Leu Glu Arg Ser Leu Gln Asn Val Ala
    65                  70                  75                  80

    Asn Tyr Leu Ile Gly Ser Leu Ala Val Thr Asp Leu Met Val Ser Val
                    85                  90              .       95

    Leu Val Leu Pro Met Ala Ala Leu Tyr Gln Val Leu Asn Lys Trp Thr
                100                 105                 110

    Leu Gly Gln Asp Ile Cys Asp Leu Phe Ile Ala Leu Asp Val Leu Cys
            115                 120                 125

    Cys Thr Ser Ser Ile Leu His Leu Cys Ala Ile Ala Leu Asp Arg Tyr
            130                 135    .            140

    Trp Ala Ile Thr Asp Pro Ile Asp Tyr Val Asn Lys Arg Thr Pro Arg
    145                 150                 155                 160

    Arg Ala Ala Val Leu Ile Ser Val Thr Trp Leu Ile Gly Phe Ser Ile
                165                 170                 175

    Ser Ile Pro Pro Met Leu Gly Trp Arg Ser Ala Glu Asp Arg Ala Asn
                180                 185                 190

    Pro Asp Ala Cys Ile Ile Ser Gln Asp Pro Gly Tyr Thr Ile Tyr Ser
            195                 200                 205
```

```
Thr Phe Gly Ala Phe Tyr Ile Pro Leu Ile Leu Met Leu Val Leu Tyr
    210                 215                 220

Gly Arg Ile Phe Lys Ala Ala Arg Phe Arg Ile Arg Lys Thr Val Lys
225                 230                 235                 240

Lys Thr Glu Lys Ala Lys Ala Ser Asp Met Cys Leu Thr Leu Ser Pro
                245                 250                 255

Ala Val Phe His Lys Arg Ala Asn Gly Asp Ala Val Ser Ala Glu Trp
                260                 265                 270

Lys Arg Gly Tyr Lys Phe Lys Pro Ser Ser Pro Cys Ala Asn Gly Ala
                275                 280                 285

Val Arg His Gly Glu Glu Met Glu Ser Leu Glu Ile Ile Glu Val Asn
    290                 295                 300

Ser Asn Ser Lys Thr His Leu Pro Leu Pro Asn Thr Pro Gln Ser Ser
305                 310                 315                 320

Ser His Glu Asn Ile Asn Glu Lys Thr Thr Gly Thr Arg Arg Lys Ile
                325                 330                 335

Ala Leu Ala Arg Glu Arg Lys Thr Val Lys Thr Leu Gly Ile Ile Met
                340                 345                 350

Gly Thr Phe Ile Phe Cys Trp Leu Pro Phe Phe Ile Val Ala Leu Val
                355                 360                 365

Leu Pro Phe Cys Ala Glu Asn Cys Tyr Met Pro Glu Trp Leu Gly Ala
    370                 375                 380

Val Ile Asn Trp Leu Gly Tyr Ser Asn Ser Leu Leu Asn Pro Ile Ile
385·                390                 395                 400

Tyr Ala Tyr Phe Asn Lys Asp Phe Gln Ser Ala Phe Lys Lys Ile Leu
                405  .               410                 415

Arg Cys Lys Phe His Arg His
                420
```

```
<210> 10
<211> 421
<212> PRT
<213> Mus musculus

<400> 10
Met Asp Met Phe Ser Leu Gly Gln Gly Asn Asn Thr Thr Thr Ser Leu
  1                 5                 10                  15

Glu Pro Phe Gly Thr Gly Gly Asn Asp Thr Gly Leu Ser Asn Val Thr
                20                  25                  30

Phe Ser Tyr Gln Val Ile Thr Ser Leu Leu Leu Gly Thr Leu Ile Phe
                35                  40                  45
```

```
Cys Ala Val Leu Gly Asn Ala Cys Val Val Ala Ala Ile Ala Leu Glu
    50              55              60

Arg Ser Leu Gln Asn Val Ala Asn Tyr Leu Ile Gly Ser Leu Ala Val
65              70              75              80 .

Thr Asp Leu Met Val Ser Val Leu Val Leu Pro Met Ala Ala Leu Tyr
            85      .       90              95

Gln Val Leu Asn Lys Trp Thr Leu Gly Gln Val Thr Cys Asp Leu Phe
            100             105             110

Ile Ala Leu Asp Val Leu Cys Cys Thr Ser Ser Ile Leu His Leu Cys
        115             120             125

Ala Ile Ala Leu Asp Arg Tyr Trp Ala Ile Thr Asp Pro Ile Asp Tyr
    130             135             140

Val Asn Lys Arg Thr Pro Arg Arg Ala Ala Ala Leu Ile Ser Leu Thr
145             150             155             160

Trp Leu Ile Gly Phe Leu Ile Ser Ile Pro Pro Met Leu Gly Trp Arg
            165             170             175

Ala Pro Glu Asp Arg Ser Asn Pro Asn Glu Cys Thr Ile Ser Lys Asp
            180             185     .       190

His Gly Tyr Thr Ile Tyr Ser Thr Phe Gly Ala Phe Tyr Ile Pro Leu
        195             200             205

Leu Leu Met Leu Val Leu Tyr Gly Arg Ile Phe Arg Ala Ala Arg Phe
    210             215             220

Arg Ile Arg Lys Thr Val Lys Lys Val Glu Lys Lys Gly Ala Gly Thr
225             230             235     ,       240

Ser Phe Gly Thr Ser Ser Ala Pro Pro Lys Lys Ser Leu Asn Gly
            245             250             255

Gln Pro Gly Ser Gly Asp Cys Arg Arg Ser Ala Glu Asn Arg Ala Val
            260             265             270

Gly Thr Pro Cys Ala Asn Gly Ala Val Arg Gln Gly Glu Asp Asp Ala
    275             280             285

Thr Leu Glu Val Ile Glu Val His Arg Val Gly Asn Ser Lys Gly Asp
    290             295             300

Leu Pro Leu Pro Ser Glu Ser Gly Ala Thr Ser Tyr Val Pro Ala Cys
305             310             315             320

Leu Glu Arg Lys Asn Glu Arg Thr Ala Glu Ala Lys Arg Lys Met Ala
            325             330             335

Leu Ala Arg Glu Arg Lys Thr Val Lys Thr Leu Gly Ile Ile Met Gly
            340             345             350

Thr Phe Ile Leu Cys Trp Leu Pro Phe Phe Ile Val Ala Leu Val Leu
    355             360             365
```

10

```
Pro Phe Cys Glu Ser Ser Cys His Met Pro Glu Leu Leu Gly Ala Ile
    370             375             380
```

```
Ile Asn Trp Leu Gly Tyr Ser Asn Ser Leu Leu Asn Pro Val Ile Tyr
385             390             395             400
```

```
Ala Tyr Phe Asn Lys Asp Phe Gln Asn Ala Phe Lys Lys Ile Ile Lys
            405             410             415
```

```
Cys Lys Phe Cys Arg
            420
```


<210> 11
<211> 423
<212> PRT
<213> Fugu rubripes

<400> 11
```
Met Asp Leu Arg Ala Thr Ser Ser Asn Asp Ser Asn Ala Thr Ser Gly
  1           5               10              15
```

```
Tyr Ser Asp Thr Ala Ala Val Asp Trp Asp Glu Gly Glu Asn Ala Thr
            20              25              30
```

```
Gly Ser Gly Ser Leu Pro Asp Pro Glu Leu Ser Tyr Gln Ile Ile Thr
        35              40              45
```

```
Ser Leu Phe Leu Gly Ala Leu Ile Leu Cys Ser Ile Phe Gly Asn Ser
    50              55              60
```

```
Cys Val Val Ala Ala Ile Ala Leu Glu Arg Ser Leu Gln Asn Val Ala
65              70              75              80
```

```
Asn Tyr Leu Ile Gly Ser Leu Ala Val Thr Asp Leu Met Val Ser Val
            85              90              95
```

```
Leu Val Leu Pro Met Ala Ala Leu Tyr Gln Val Leu Asn Lys Trp Thr
            100             105             110
```

```
Leu Gly Gln Asp Ile Cys Asp Leu Phe Ile Ala Leu Asp Val Leu Cys
        115             120             125
```

```
Cys Thr Ser Ser Ile Leu His Leu Cys Ala Ile Ala Leu Asp Arg Tyr
    130             135             140
```

```
Trp Ala Ile Thr Asp Pro Ile Asp Tyr Val Asn Lys Arg Thr Pro Arg
145             150             155             160
```

```
Arg Ala Ala Val Leu Ile Ser Val Thr Trp Leu Ile Gly Phe Ser Ile
            165             170             175
```

```
Ser Ile Pro Pro Met Leu Gly Trp Arg Ser Ala Glu Asp Arg Ala Asn
        180             185             190
```

```
Pro Asp Ala Cys Ile Ile Ser Gln Asp Pro Gly Tyr Thr Ile Tyr Ser
        195             200             205
```

```
Thr Phe Gly Ala Phe Tyr Ile Pro Leu Ile Leu Met Leu Val Leu Tyr
    210                 215                 220

Gly Arg Ile Phe Lys Ala Ala Arg Phe Arg Ile Arg Lys Thr Val Lys
225                 230                 235                 240

Lys Thr Glu Lys Ala Lys Ala Ser Asp Met Cys Leu Thr Leu Ser Pro
            245                 250                 255

Ala Val Phe His Lys Arg Ala Asn Gly Asp Ala Val Ser Ala Glu Trp
            260                 265                 270

Lys Arg Gly Tyr Lys Phe Lys Pro Ser Ser Pro Cys Ala Asn Gly Ala
        275                 280                 285

Val Arg His Gly Glu Glu Met Glu Ser Leu Glu Ile Ile Glu Val Asn
    290                 295                 300

Ser Asn Ser Lys Thr His Leu Pro Leu Pro Asn Thr Pro Gln Ser Ser
305                 310                 315                 320

Ser His Glu Asn Ile Asn Glu Lys Thr Thr Gly Thr Arg Arg Lys Ile
            325                 330                 335

Ala Leu Ala Arg Glu Arg Lys Thr Val Lys Thr Leu Gly Ile Ile Met
            340                 345                 350

Gly Thr Phe Ile Phe Cys Trp Leu Pro Phe Phe Ile Val Ala Leu Val
        355                 360                 365

Leu Pro Phe Cys Ala Glu Asn Cys Tyr Met Pro Glu Trp Leu Gly Ala
    370                 375                 380

Val Ile Asn Trp Leu Gly Tyr Ser Asn Ser Leu Leu Asn Pro Ile Ile
385                 390                 395                 400

Tyr Ala Tyr Phe Asn Lys Asp Phe Gln Ser Ala Phe Lys Lys Ile Leu
            405                 410                 415

Arg Cys Lys Phe His Arg His
            420
```

```
<210> 12
<211> 509
<212> PRT
<213> Lymnaea stagnalis

<400> 12
Met Ala Asn Phe Thr Phe Gly Asp Leu Ala Leu Asp Val Ala Arg Met
  1             5                 10                  15

Gly Gly Leu Ala Ser Thr Pro Ser Gly Leu Arg Ser Thr Gly Leu Thr
            20                  25                  30

Thr Pro Gly Leu Ser Pro Thr Gly Leu Val Thr Ser Asp Phe Asn Asp
        35                  40                  45

Ser Tyr Gly Leu Thr Gly Gln Phe Ile Asn Gly Ser His Ser Ser Arg
```

```
            50                      55                      60

  Ser Arg Asp Asn Ala Ser Ala Asn Asp Thr Ser Ala Thr Asn Met Thr
  65                  70                  75                  80

  Asp Asp Arg Tyr Trp Ser Leu Thr Val Tyr Ser His Glu His Leu Val
                  85                  90                      95

  Leu Thr Ser Val Ile Leu Gly Leu Phe Val Leu Cys Cys Ile Ile Gly
                 100             105             110

  Asn Cys Phe Val Ile Ala Ala Val Met Leu Glu Arg Ser Leu His Asn
             115             120             125

  Val Ala Asn Tyr Leu Ile Leu Ser Leu Ala Val Ala Asp Leu Met Val
     130             135             140

  Ala Val Leu Val Met Pro Leu Ser Val Val Ser Glu Ile Ser Lys Val
  145             150             155             160

  Trp Phe Leu His Ser Glu Val Cys Asp Met Trp Ile Ser Val Asp Val
                 165             170             175

  Leu Cys Cys Thr Ala Ser Ile Leu His Leu Val Ala Ile Ala Met Asp
             180             185             190

  Arg Tyr Trp Ala Val Thr Ser Ile Asp Tyr Ile Arg Arg Arg Ser Ala
     195             200             205

  Arg Arg Ile Leu Leu Met Ile Met Val Val Trp Ile Val Ala Leu Phe
     210             215             220

  Ile Ser Ile Pro Pro Leu Phe Gly Trp Arg Asp Pro Asn Asn Asp Pro
  225             230             235             240

  Asp Lys Thr Gly Thr Cys Ile Ile Ser Gln Asp Lys Gly Tyr Thr Ile
             245             250             255

  Phe Ser Thr Val Gly Ala Phe Tyr Leu Pro Met Leu Val Met Met Ile
             260             265             270

  Ile Tyr Ile Arg Ile Trp Leu Val Ala Arg Ser Arg Ile Arg Lys Asp
     275             280             285

  Lys Phe Gln Met Thr Lys Ala Arg Leu Lys Thr Glu Glu Thr Thr Leu
     290             295             300

  Val Ala Ser Pro Lys Thr Glu Tyr Ser Val Val Ser Asp Cys Asn Gly
  305             310             315             320

  Cys Asn Ser Pro Asp Ser Thr Thr Glu Lys Lys Lys Arg Arg Ala Pro
                 325             330             335

  Phe Lys Ser Tyr Gly Cys Ser Pro Arg Pro Glu Arg Lys Lys Asn Arg
             340             345             350

  Ala Lys Lys Leu Pro Glu Asn Ala Asn Gly Val Asn Ser Asn Ser Ser
     355             360             365
```

```
Ser Ser Glu Arg Leu Lys Gln Ile Gln Ile Glu Thr Ala Glu Ala Phe
    370             375             380
```

```
Ala Asn Gly Cys Ala Glu Glu Ala Ser Ile Ala Met Leu Glu Arg Gln
385             390             395                 400
```

```
Cys Asn Asn Gly Lys Lys Ile Ser Ser Asn Asp Thr Pro Tyr Ser Arg
            405             410             415
```

```
Thr Arg Glu Lys Leu Glu Leu Lys Arg Glu Arg Lys Ala Ala Arg Thr
        420             425             430
```

```
Leu Ala Ile Ile Thr Gly Ala Phe Leu Ile Cys Trp Leu Pro Phe Phe
        435             440             445
```

```
Ile Ile Ala Leu Ile Gly Pro Phe Val Asp Pro Glu Gly Ile Pro Pro
    450             455             460
```

```
Phe Ala Arg Ser Phe Val Leu Trp Leu Gly Tyr Phe Asn Ser Leu Leu
465             470             475             480
```

```
Asn Pro Ile Ile Tyr Thr Ile Phe Ser Pro Glu Phe Arg Ser Ala Phe
            485             490             495
```

```
Gln Lys Ile Leu Phe Gly Lys Tyr Arg Arg Gly His Arg
            500             505
```

```
<210> 13
<211> 572
<212> PRT
<213> Homo sapiens
```

```
<400> 13
Met Thr Phe Arg Asp Leu Leu Ser Val Ser Phe Glu Gly Pro Arg Pro
1               5               10              15
```

```
Asp Ser Ser Ala Gly Gly Ser Ser Ala Gly Gly Gly Gly Ser Ala
            20              25              30
```

```
Gly Gly Ala Ala Pro Ser Glu Gly Pro Ala Val Gly Gly Val Pro Gly
        35              40              45
```

```
Gly Ala Gly Gly Gly Gly Gly Val Val Gly Ala Gly Ser Gly Glu Asp
    50              55              60
```

```
Asn Arg Ser Ser Ala Gly Glu Pro Gly Ser Ala Gly Ala Gly Gly Asp
65              70              75              80
```

```
Val Asn Gly Thr Ala Ala Val Gly Gly Leu Val Val Ser Ala Gln Gly
            85              90              95
```

```
Val Gly Val Gly Val Phe Leu Ala Ala Phe Ile Leu Met Ala Val Ala
        100             105             110
```

```
Gly Asn Leu Leu Val Ile Leu Ser Val Ala Cys Asn Arg His Leu Gln
    115             120             125
```

```
Thr Val Thr Asn Tyr Phe Ile Val Asn Leu Ala Val Ala Asp Leu Leu
```

```
             130                    135                    140
Leu Ser Ala Thr Val Leu Pro Phe Ser Ala Thr Met Glu Val Leu Gly
145                 150                 155                 160

Phe Trp Ala Phe Gly Arg Ala Phe Cys Asp Val Trp Ala Ala Val Asp
                165                 170                 175

Val Leu Cys Cys Thr Ala Ser Ile Leu Ser Leu Cys Thr Ile Ser Val
            180                 185                 190

Asp Arg Tyr Val Gly Val Arg His Ser Leu Lys Tyr Pro Ala Ile Met
        195                 200                 205

Thr Glu Arg Lys Ala Ala Ala Ile Leu Ala Leu Leu Trp Val Val Ala
    210                 215                 220

Leu Val Val Ser Val Gly Pro Leu Leu Gly Trp Lys Glu Pro Val Pro
225                 230                 235                 240

Pro Asp Glu Arg Phe Cys Gly Ile Thr Glu Glu Ala Gly Tyr Ala Val
                245                 250                 255

Phe Ser Ser Val Cys Ser Phe Tyr Leu Pro Met Ala Val Ile Val Val
            260                 265                 270

Met Tyr Cys Arg Val Tyr Val Val Ala Arg Ser Thr Thr Arg Ser Leu
        275                 280                 285

Glu Ala Gly Val Lys Arg Glu Arg Gly Lys Ala Ser Glu Val Val Leu
    290                 295                 300

Arg Ile His Cys Arg Gly Ala Ala Thr Gly Ala Asp Gly Ala His Gly
305                 310                 315                 320

Met Arg Ser Ala Lys Gly His Thr Phe Arg Ser Ser Leu Ser Val Arg
                325                 330                 335

Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys Thr Leu Ala Ile
            340                 345                 350

Val Val Gly Val Phe Val Leu Cys Trp Phe Pro Phe Phe Phe Val Leu
            355                 360                 365

Pro Leu Gly Ser Leu Phe Pro Gln Leu Lys Pro Ser Glu Gly Val Phe
370                 375                 380

Lys Val Ile Phe Trp Leu Gly Tyr Phe Asn Ser Cys Val Asn Pro Leu
385                 390                 395                 400

Ile Tyr Pro Cys Ser Ser Arg Glu Phe Lys Arg Ala Phe Leu Arg Leu
            405                 410                 415

Leu Arg Cys Gln Cys Arg Arg Arg Arg Arg Arg Arg Pro Leu Trp Arg
            420                 425                 430

Val Tyr Gly His His Trp Arg Ala Ser Thr Ser Gly Leu Arg Gln Asp
435                 440                 445
```

15

```
Cys Ala Pro Ser Ser Gly Asp Ala Pro Pro Gly Ala Pro Leu Ala Leu
    450             455             460
Thr Ala Leu Pro Asp Pro Asp Pro Glu Pro Pro Gly Thr Pro Glu Met
465             470             475             480
Gln Ala Pro Val Ala Ser Arg Arg Lys Pro Pro Ser Ala Phe Arg Glu
            485             490             495
Trp Arg Leu Leu Gly Pro Phe Arg Arg Pro Thr Thr Gln Leu Arg Ala
            500             505             510
Lys Val Ser Ser Leu Ser His Lys Ile Arg Ala Gly Gly Ala Gln Arg
            515             520             525
Ala Glu Ala Ala Cys Ala Gln Arg Ser Glu Val Glu Ala Val Ser Leu
    530             535             540
Gly Val Pro His Glu Val Ala Glu Gly Ala Thr Cys Gln Ala Tyr Glu
545             550             555             560
Leu Ala Asp Tyr Ser Asn Leu Arg Glu Thr Asp Ile
            565             570
```

```
<210> 14
<211> 562
<212> PRT
<213> Mus musculus

<400> 14
Met Thr Phe Arg Asp Ile Leu Ser Val Thr Phe Glu Gly Pro Arg Ala
  1             5              10              15
Ser Ser Ser Thr Gly Gly Ser Gly Ala Gly Gly Ala Gly Thr Val
            20              25              30
Gly Pro Glu Gly Pro Ala Val Gly Gly Val Pro Gly Ala Thr Gly Gly
        35              40              45
Ser Ala Val Val Gly Thr Gly Ser Gly Glu Asp Asn Gln Ser Ser Thr
    50              55              60
Ala Glu Ala Gly Ala Ala Ala Ser Gly Glu Val Asn Gly Ser Ala Ala
65              70              75              80
Val Gly Gly Leu Val Val Ser Ala Gln Gly Val Gly Val Gly Val Phe
            85              90              95
Leu Ala Ala Phe Ile Leu Thr Ala Val Ala Gly Asn Leu Leu Val Ile
            100             105             110
Leu Ser Val Ala Cys Asn Arg His Leu Gln Thr Val Thr Asn Tyr Phe
        115             120             125
Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Ser Ala Ala Val Leu
    130             135             140
Pro Phe Ser Ala Thr Met Glu Val Leu Gly Phe Trp Pro Phe Gly Arg
```

```
      145                150                155                160

        Thr Phe Cys Asp Val Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
                    165                170                175

        Ser Ile Leu Ser Leu Cys Thr Ile Ser Val Asp Arg Tyr Val Gly Val
                    180                185                190

        Arg His Ser Leu Lys Tyr Pro Ala Ile Met Thr Glu Arg Lys Ala Ala
                195                200                205

        Ala Ile Leu Ala Leu Leu Trp Ala Val Ala Leu Val Val Ser Val Gly
        210                215                220

        Pro Leu Leu Gly Trp Lys Glu Pro Val Pro Pro Asp Glu Arg Phe Cys
        225                230                235                240

        Gly Ile Thr Glu Glu Val Gly Tyr Ala Ile Phe Ser Ser Val Cys Ser
                    245                250                255

        Phe Tyr Leu Pro Met Ala Val Ile Val Val Met Tyr Cys Arg Val Tyr
                    260                265                270

        Val Val Ala Arg Ser Thr Thr Arg Ser Leu Glu Ala Gly Ile Lys Arg
                275                280                285

        Glu Pro Gly Lys Ala Ser Glu Val Val Leu Arg Ile His Cys Arg Gly
            290                295                300

        Ala Ala Thr Ser Ala Lys Gly Asn Pro Gly Thr Gln Ser Ser Lys Gly
        305                310                315                320

        His Thr Leu Arg Ser Ser Leu Ser Val Arg Leu Leu Lys Phe Ser Arg
                    325                330                335

        Glu Lys Lys Ala Ala Lys Thr Leu Ala Ile Val Val Gly Val Phe Val
                340                345                350

        Leu Cys Trp Phe Pro Phe Phe Phe Val Leu Pro Leu Gly Ser Leu Phe
                355                360                365

        Pro Gln Leu Lys Pro Ser Glu Gly Val Phe Lys Val Ile Phe Trp Leu
        370                375                380

        Gly Tyr Phe Asn Ser Cys Val Asn Pro Leu Ile Tyr Pro Cys Ser Ser
        385                390                395                400

        Arg Glu Phe Lys Arg Ala Phe Leu Arg Leu Leu Arg Cys Gln Cys Arg
                405                410                415

        Arg Arg Arg Arg Arg Leu Trp Pro Ser Leu Arg Pro Pro Leu Ala Ser
                420                425                430

        Leu Asp Arg Arg Pro Ala Leu Arg Leu Cys Pro Gln Pro Ala His Arg
                435                440                445

        Thr Pro Arg Gly Ser Pro Ser Pro His Cys Thr Pro Arg Pro Gly Leu
        450                455                460
```

17

```
Arg Arg His Ala Gly Gly Ala Gly Phe Gly Leu Arg Pro Ser Lys Ala
465             470             475             480

Ser Leu Arg Leu Arg Glu Trp Arg Leu Leu Gly Pro Leu Gln Arg Pro
                485             490             495

Thr Thr Gln Leu Arg Ala Lys Val Ser Ser Leu Ser His Lys Phe Arg
            500             505             510

Ser Gly Gly Ala Arg Arg Ala Glu Thr Ala Cys Ala Leu Arg Ser Glu
        515             520             525

Val Glu Ala Val Ser Leu Asn Val Pro Gln Asp Gly Ala Glu Ala Val
    530             535             540

Ile Cys Gln Ala Tyr Glu Pro Gly Asp Leu Ser Asn Leu Arg Glu Thr
545             550             555             560

Asp Ile
```

```
<210> 15
<211> 499
<212> PRT
<213> Homo sapiens

<400> 15
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr Gln
 1           5               10                  15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20              25                  30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly Asn Ile Leu Val Ile
        35              40                  45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50              55                  60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65              70              75                  80

Pro Phe Ser Ala Ile Phe Glu Val Leu Gly Tyr Trp Ala Phe Gly Arg
            85              90                  95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
            100             105             110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
        115             120             125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Leu
    130             135             140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145             150             155             160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
```

18

```
                 165                    170                    175
     Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
                 180                    185                    190

     Phe Tyr Leu Pro Leu Ala Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
                 195                    200                    205

     Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
         210                    215                    220

     Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
     225                    230                    235                    240

     Ala Pro Ala Gly Gly Ser Gly Met Ala Ser Ala Lys Thr Lys Thr His
                     245                    250                    255

     Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
                 260                    265                    270

     Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
                 275                    280                    285

     Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Ser
         290                    295                    300

     Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
     305                    310                    315                    320

     Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
                     325                    330                    335

     Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Arg Arg Lys Gln Ser Ser
                 340                    345                    350

     Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Val Glu
                 355                    360                    365

     Gly Gln His Lys Asp Met Val Arg Ile Pro Val Gly Ser Arg Glu Thr
         370                    375                    380

     Phe Tyr Arg Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
     385                    390                    395                    400

     Ser Ser Met Pro Arg Gly Ser Ala Arg Ile Thr Val Ser Lys Asp Gln
                     405                    410                    415

     Ser Ser Cys Thr Thr Ala Arg Thr Lys Ser Arg Ser Val Thr Arg Leu
                 420                    425                    430

     Glu Cys Ser Gly Met Ile Leu Ala His Cys Asn Leu Arg Leu Pro Gly
             435                    440                    445

     Ser Arg Asp Ser Pro Ala Ser Ala Ser Gln Ala Ala Gly Thr Thr Gly
         450                    455                    460

     Asp Val Pro Pro Gly Arg Arg His Gln Ala Gln Leu Ile Phe Val Phe
     465                    470                    475                    480
```

```
Leu Val Glu Thr Gly Phe His His Val Gly Gln Asp Asp Leu Asp Leu
                485                  490                  495

Leu Thr Ser



<210> 16
<211> 429
<212> PRT
<213> Homo sapiens

<400> 16
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr Gln
  1               5                  10                  15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
              20                  25                  30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly Asn Ile Leu Val Ile
          35                  40                  45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
      50                  55                  60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
 65              70                  75                      80

Pro Phe Ser Ala Ile Phe Glu Val Leu Gly Tyr Trp Ala Phe Gly Arg
              85                  90                  95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
             100                 105                 110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
         115                 120                 125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Leu
     130                 135                 140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145                 150                 155                 160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
             165                 170                 175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
             180                 185                 190

Phe Tyr Leu Pro Leu Ala Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
         195                 200                 205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
     210                 215                 220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225                 230                 235                 240

Ala Pro Ala Gly Gly Ser Gly Met Ala Ser Ala Lys Thr Lys Thr His
```

                          245   250   255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
                260   265   270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
            275   280   285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Ser
        290   295   300

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305   310   315   320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
                325   330   335

Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Arg Arg Lys Gln Ser Ser
            340   345   350

Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Val Glu
        355   360   365

Gly Gln His Lys Asp Met Val Arg Ile Pro Val Gly Ser Arg Glu Thr
    370   375   380

Phe Tyr Arg Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
385   390   395   400

Ser Ser Met Pro Arg Gly Ser Ala Arg Ile Thr Val Ser Lys Asp Gln
                405   410   415

Ser Ser Cys Thr Thr Ala Arg Gly His Thr Pro Met Thr
            420   425

<210> 17
<211> 455
<212> PRT
<213> Homo sapiens

<400> 17
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr Gln
  1   5   10   15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20   25   30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly Asn Ile Leu Val Ile
        35   40   45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50   55   60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65   70   75   80

Pro Phe Ser Ala Ile Phe Glu Val Leu Gly Tyr Trp Ala Phe Gly Arg
                85   90   95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
         100   105   110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
         115   120   125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Leu
    130   135   140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145   150   155   160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
             165   170   175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
         180   185   190

Phe Tyr Leu Pro Leu Ala Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
    195   200   205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
    210   215   220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225   230   235   240

Ala Pro Ala Gly Gly Ser Gly Met Ala Ser Ala Lys Thr Lys Thr His
             245   250   255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
             260   265   270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
    275   280   285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Ser
    290   295   300

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305   310   315   320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
             325   330   335

Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Cys Arg Lys Gln Ser Ser
             340   345   350

Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Val Glu
         355   360   365

Gly Gln His Lys Asp Met Val Arg Ile Pro Val Gly Ser Arg Glu Thr
    370   375   380

Phe Tyr Arg Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
385   390   395   400

Ser Ser Met Pro Arg Gly Ser Ala Arg Ile Thr Val Ser Lys Asp Gln

```
                        405                     410                     415
Ser Ser Cys Thr Thr Ala Arg Arg Gly Met Asp Cys Arg Tyr Phe Thr
            420                     425                 430

Lys Asn Cys Arg Glu His Ile Lys His Val Asn Phe Met Met Pro Pro
        435                     440           .         445

Trp Arg Lys Gly Leu Glu Cys
    450                 455
```

```
<210> 18
<211> 466
<212> PRT
<213> Rattus norvegicus

<400> 18
Met Val Leu Leu Ser Glu Asn Ala Ser Glu Gly Ser Asn Cys Thr His
  1             5                  10                  15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20                  25                  30

Leu Gly Gly Leu Ile Ile Phe Gly Val Leu Gly Asn Ile Leu Val Ile
        35                  40                  45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50                  55                  60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65                  70                  75                  80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Gly Tyr Trp Ala Phe Gly Arg
            85                  90                  95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
        100                 105                 110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
    115                 120                 125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Val
    130                 135                 140

Arg Ala Leu Leu Cys Val Trp Val Leu Ser Leu Val Ile Ser Ile Gly
.145                 150                 155                 160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
            165 `                 170                 175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
        180                 185                 190

Phe Tyr Val Pro Leu Ala Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
    195                 200                 205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
    210                 215                 220
```

```
Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225                 230                 235                 240

Val Pro Ala Glu Gly Gly Val Ser Ser Ala Lys Asn Lys Thr His
                245                 250                 255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
                260                 265                 270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
                275                 280                 285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Ser
        290                 295                 300

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305                 310                 315                 320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
                325                 330                 335

Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Arg Arg Arg Gln Ser Ser
            340                 345                 350

Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Leu Glu
        355                 360                 365

Gly Gln His Arg Asp Met Val Arg Ile Pro Val Gly Ser Gly Glu Thr
    370                 375                 380

Phe Tyr Lys Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
385                 390                 395                 400

Ser Ser Met Pro Gln Gly Ser Ala Arg Ile Thr Val Pro Lys Asp Gln
                405                 410                 415

Ser Ala Cys Thr Thr Ala Arg Val Arg Ser Lys Ser Phe Leu Gln Val
                420                 425                 430

Cys Cys Cys Val Gly Ser Ser Ala Pro Arg Pro Glu Glu Asn His Gln
        435                 440                 445

Val Pro Thr Ile Lys Ile His Thr Ile Ser Leu Gly Glu Asn Gly Glu
    450                 455                 460

Glu Val
465


<210> 19
<211> 466
<212> PRT
<213> Mus musculus

<400> 19
Met Val Leu Leu Ser Glu Asn Ala Ser Glu Gly Ser Asn Cys Thr His
  1                 5                  10                  15
```

24

Pro Pro Ala Gln Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20   25   30

Leu Gly Gly Leu Ile Ile Phe Gly Val Leu Gly Asn Ile Leu Val Ile
            35   40   45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
         50   55   60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65   70   75   .   80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Gly Tyr Trp Ala Phe Gly Arg
                  85   90   95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
               100   105   110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
            115   120   125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Val
   130   135   140

Arg Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145   150   155   160

Pro Leu Phe Gly Trp Arg Gln Gln Ala Pro Glu Asp Glu Thr Ile Cys
               165   170   175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
            180   185   190

Phe Tyr Val Pro Leu Thr Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
      195   200   205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
   210   215   220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225   230   235   240

Val Pro Ala Glu Gly Ser Gly Val Ser Ser Ala Lys Asn Lys Thr His
               245   ·250   255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
            260   265   270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
         275   280   285   .

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asn Phe Lys Pro Pro
   290   295   300
                           .

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305   310   315   320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
               325   330   335

25

.

```
Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Arg Arg Arg Gln Ser Ser
        340             345             350

Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Val Glu
        355             360             365

Glu Gln His Arg Gly Met Val Arg Ile Pro Val Gly Ser Gly Glu Thr
    370             375             380

Phe Tyr Lys Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
385             390             395             400

Ser Ser Met Pro Gln Gly Ser Ala Arg Ile Thr Met Pro Lys Asp Gln
            405             410             415

Ser Ala Cys Thr Thr Ala Arg Val Arg Ser Lys Ser Phe Leu Gln Val
            420             425             430

Cys Cys Cys Val Gly Ser Ser Thr Pro Arg Pro Glu Glu Asn His Gln
        435             440             445

Val Pro Thr Ile Lys Ile His Thr Ile Ser Leu Gly Glu Asn Gly Glu
    450             455             460

Glu Val
465


<210> 20
<211> 466
<212> PRT
<213> Bos taurus

<400> 20
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr His
  1             5               10              15

Pro Pro Pro Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20              25              30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly Asn Ile Leu Val Ile
        35              40              45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50              55              60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65              70              75              80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Gly Tyr Trp Ala Phe Gly Arg
                85              90              95

Val Phe Cys Asn Val Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
        100             105             110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
        115             120             125
```

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Lys Arg Gly Leu
    130   135   140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145   150   155   160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
            165   170   175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
        180   185   190

Phe Tyr Val Pro Leu Thr Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
        195   200   205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
    210   215   220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225   230   235   240

Ala Gln Val Gly Gly Ser Gly Val Thr Ser Ala Lys Asn Lys Thr His
            245   250   255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
        260   265   270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
    275   280   285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Arg Pro Ser
    290   295   300

Glu Thr Val Phe Lys Ile Ala Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305   310   315   320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
            325   330   335

Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Arg Arg Lys Gln Ser Ser
        340   345   350

Lys His Thr Leu Gly Tyr Thr Leu His Ala Pro Ser His Val Leu Glu
    355   360   365

Gly Gln His Lys Asp Leu Val Arg Ile Pro Val Gly Ser Ala Glu Thr
    370   375   380

Phe Tyr Lys Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Ile Phe
385   390   395   400

Ser Ser Leu Pro Arg Gly Ser Ala Arg Met Ala Val Ala Arg Asp Pro
            405   410   415

Ser Ala Cys Thr Thr Ala Arg Val Arg Ser Lys Ser Phe Leu Gln Val
            420   425   430

Cys Cys Cys Leu Gly Pro Ser Thr Pro Ser His Gly Glu Asn His Gln
            435   440   445

27

Ile Pro Thr Ile Lys Ile His Thr Ile Ser Leu Ser Glu Asn Gly Glu
    450   455   460

Glu Val
465

<210> 21
<211> 295
<212> PRT
<213> Canis familiaris

<400> 21
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr His
  1   5   10   15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20   25   30

Leu Gly Gly Leu Ile Ile Phe Gly Val Leu Gly Asn Ile Leu Val Ile
        35   40   45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50   55   60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
 65   70   75   80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Gly Tyr Trp Ala Phe Gly Arg
            85   90   95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
            100   105   110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
        115   120   125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Lys Arg Gly Leu
    130   135   140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145   150   155   160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
            165   170   175

Gln Ile Thr Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
            180   185   190

Phe Tyr Val Pro Leu Thr Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
        195   200   205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
    210   215   220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225   230   235   240

```
Ala Pro Val Gly Gly Thr Gly Val Ser Ser Ala Lys Asn Lys Thr His
              245                 250                 255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
              260                 265                 270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
              275                 280                 285

Phe Leu Val Met Pro Ile Gly
              290                 295


<210> 22
<211> 466
<212> PRT
<213> Oryctolagus cuniculus

<400> 22
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr His
  1           5                  10                  15

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
              20                  25                  30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly Asn Ile Leu Val Ile
              35                  40                  45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
      50                  55                  60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
 65                  70                  75                  80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Gly Tyr Trp Ala Phe Gly Arg
                  85                  90                  95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
              100                 105                 110

Ser Ile Ile Ser Leu Cys Val Ile Ser Ile Asp Arg Tyr Ile Gly Val
              115                 120                 125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Leu
      130                 135                 140

Arg Ala Leu Leu Cys Val Trp Ala Phe Ser Leu Val Ile Ser Val Gly
145                 150                 155                 160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Asp Asp Glu Thr Ile Cys
              165                 170                 175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
              180                 185                 190

Phe Tyr Val Pro Leu Thr Ile Ile Leu Ala Met Tyr Cys Arg Val Tyr
      195                 200                 205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
```

```
        210                        215                        220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225             230             235                 240

Ala Pro Ala Gly Gly Ser Gly Val Ala Ser Ala Lys Asn Lys Thr His
            245             250                 255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
            260             265                 270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
            275             280             285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Pro
    290             295             300

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305             310             315                 320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala
            325             330             335

Phe Gln Asn Val Leu Lys Ile Gln Cys Leu Arg Arg Lys Gln Ser Ser
            340             345             350

Lys His Ala Leu Gly Tyr Thr Leu His Ala Pro Ser Gln Ala Leu Glu
    355             360             365

Gly Gln His Lys Asp Met Val Arg Ile Pro Val Gly Ser Gly Glu Thr
    370             375             380

Phe Tyr Lys Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
385             390             395                 400

Ser Ser Met Pro Arg Gly Ser Ala Arg Ile Thr Val Pro Lys Asp Gln
            405             410             415

Ser Ala Cys Thr Thr Ala Arg Val Arg Ser Lys Ser Phe Leu Gln Val
            420             425             430

Cys Cys Cys Val Gly Pro Ser Thr Pro Asn Pro Gly Glu Asn His Gln
    435             440             445

Val Pro Thr Ile Lys Ile His Thr Ile Ser Leu Ser Glu Asn Gly Glu
    450             455             460

Glu Val
465


<210> 23
<211> 466
<212> PRT
<213> Homo sapiens

<400> 23
Met Val Phe Leu Ser Gly Asn Ala Ser Asp Ser Ser Asn Cys Thr Gln
  1             5                 10              15
```

Pro Pro Ala Pro Val Asn Ile Ser Lys Ala Ile Leu Leu Gly Val Ile
            20   25   30

Leu Gly Gly Leu Ile Leu Phe Gly Val Leu Gly·Asn Ile Leu Val Ile
        35   40   45

Leu Ser Val Ala Cys His Arg His Leu His Ser Val Thr His Tyr Tyr
    50   55   60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Thr Ser Thr Val Leu
65   70   75   80

Pro Phe Ser Ala Ile Phe Glu Val Leu Gly Tyr Trp Ala Phe Gly Arg
                85   90   95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
            100   105   110

Ser Ile Met Gly Leu Cys Ile Ile Ser Ile Asp Arg Tyr Ile Gly Val
        115   120   125

Ser Tyr Pro Leu Arg Tyr Pro Thr Ile Val Thr Gln Arg Arg Gly Leu
    130   135   140

Met Ala Leu Leu Cys Val Trp Ala Leu Ser Leu Val Ile Ser Ile Gly
145   150   155   160

Pro Leu Phe Gly Trp Arg Gln Pro Ala Pro Glu Asp Glu Thr Ile Cys
                165   170   175

Gln Ile Asn Glu Glu Pro Gly Tyr Val Leu Phe Ser Ala Leu Gly Ser
            180   .   185   190

Phe Tyr Leu Pro Leu Ala Ile Ile Leu Val Met Tyr Cys Arg Val Tyr
    195   200   205

Val Val Ala Lys Arg Glu Ser Arg Gly Leu Lys Ser Gly Leu Lys Thr
    210   215   220

Asp Lys Ser Asp Ser Glu Gln Val Thr Leu Arg Ile His Arg Lys Asn
225   230   235   240

Ala Pro Ala Gly Gly Ser Gly Met Ala Ser Ala Lys Thr Lys Thr His
            245   250   255

Phe Ser Val Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys
            260   265   270

Thr Leu Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe
        275   280   285

Phe Leu Val Met Pro Ile Gly Ser Phe Phe Pro Asp Phe Lys Pro Ser
    290   295   300

Glu Thr Val Phe Lys Ile Val Phe Trp Leu Gly Tyr Leu Asn Ser Cys
305   310   315   320

Ile Asn Pro Ile Ile Tyr Pro Cys Ser Ser Gln Glu Phe Lys Lys Ala

31

```
                     325                    330                    335

      Phe Gln Asn Val Leu Arg Ile Gln Cys Leu Cys Arg Lys Gln Ser Ser
                 340                 345                 350

      Lys His Ala Leu Gly Tyr Thr Leu His Pro Pro Ser Gln Ala Val Glu
                 355                 360                 365

      Gly Gln His Lys Asp Met Val Arg Ile Pro Val Gly Ser Arg Glu Thr
                 370                 375                 380

      Phe Tyr Arg Ile Ser Lys Thr Asp Gly Val Cys Glu Trp Lys Phe Phe
      385                 390                 395                 400

      Ser Ser Met Pro Arg Gly Ser Ala Arg Ile Thr Val Ser Lys Asp Gln
                 405                 410                 415

      Ser Ser Cys Thr Thr Ala Arg Val Arg Ser Lys Ser Phe Leu Gln Val
                 420                 425                 430

      Cys Cys Cys Val Gly Pro Ser Thr Pro Ser Leu Asp Lys Asn His Gln
                 435                 440                 445

      Val Pro Thr Ile Lys Val His Thr Ile Ser Leu Ser Glu Asn Gly Glu
                 450                 455                 460

      Glu Val
      465
```

```
<210> 24
<211> 470
<212> PRT
<213> Oryzias latipes

<400> 24
Met Thr Pro Ser Ser Val Thr Leu Asn Cys Ser Asn Cys Ser His Val
  1               5                  10                  15

Leu Ala Pro Glu Leu Asn Thr Val Lys Ala Val Val Leu Gly Met Val
             20                  25                  30

Leu Gly Ile Phe Ile Leu Phe Gly Val Ile Gly Asn Ile Leu Val Ile
             35                  40                  45

Leu Ser Val Val Cys His Arg His Leu Gln Thr Val Thr Tyr Tyr Phe
        50                  55                  60

Ile Val Asn Leu Ala Val Ala Asp Leu Leu Leu Ser Ser Thr Val Leu
 65                  70                  75                  80

Pro Phe Ser Ala Ile Phe Glu Ile Leu Asp Arg Trp Val Phe Gly Arg
                 85                  90                  95

Val Phe Cys Asn Ile Trp Ala Ala Val Asp Val Leu Cys Cys Thr Ala
                100                 105                 110

Ser Ile Met Ser Leu Cys Val Ile Ser Val Asp Arg Tyr Ile Gly Val
                115                 120                 125
```

```
Ser Tyr Pro Leu Arg Tyr Pro Ala Ile Met Thr Lys Arg Arg Ala Leu
    130             135             140

Leu Ala Val Met Leu Leu Trp Val Leu Ser Val Ile Ile Ser Ile Gly
145             150             155             160

Pro Leu Phe Gly Trp Lys Glu Pro Ala Pro Glu Asp Glu Thr Val Cys
            165             170             175

Lys Ile Thr Glu Glu Pro Gly Tyr Ala Ile Phe Ser Ala Val Gly Ser
            180             185             190

Phe Tyr Leu Pro Leu Ala Ile Ile Leu Ala Met Tyr Cys Arg Val Tyr
        195             200             205

Val Val Ala Gln Lys Glu Ser Arg Gly Leu Lys Glu Gly Gln Lys Ile
    210             215             220

Glu Lys Ser Asp Ser Glu Gln Val Ile Leu Arg Met His Arg Gly Asn
225             230             235             240

Thr Thr Val Ser Glu Asp Glu Ala Leu Arg Ser Arg Thr His Phe Ala
            245             250             255

Leu Arg Leu Leu Lys Phe Ser Arg Glu Lys Lys Ala Ala Lys Thr Leu
        260             265             270

Gly Ile Val Val Gly Cys Phe Val Leu Cys Trp Leu Pro Phe Phe Leu
    275             280             285

Val Leu Pro Ile Gly Ser Ile Phe Pro Ala Tyr Arg Pro Ser Asp Thr
290             295             300

Val Phe Lys Ile Thr Phe Trp Leu Gly Tyr Phe Asn Ser Cys Ile Asn
305             310             315             320

Pro Ile Ile Tyr Leu Cys Ser Asn Gln Glu Phe Lys Lys Ala Phe Gln
            325             330             335

Ser Leu Leu Gly Val His Cys Leu Arg Met Thr Pro Arg Ala His His
        340             345             350

His His Leu Ser Val Gly Gln Ser Gln Thr Gln Gly His Ser Leu Thr
    355             360             365

Ile Ser Leu Asp Ser Lys Gly Ala Pro Cys Arg Leu Ser Pro Ser Ser
370             375             380

Ser Val Ala Leu Ser Arg Thr Pro Ser Ser Arg Asp Ser Arg Glu Trp
385             390             395             400

Arg Val Phe Ser Gly Gly Pro Ile Asn Ser Gly Pro Gly Pro Thr Glu
            405             410             415

Ala Gly Arg Ala Lys Val Ala Lys Leu Cys Asn Lys Ser Leu His Arg
        420             425             430

Thr Cys Cys Cys Ile Leu Arg Ala Arg Thr Pro Thr Gln Asp Pro Ala
```

```
        435                    440                    445
```

Pro Leu Gly Asp Leu Pro Thr Ile Lys Ile His Gln Leu Ser Leu Ser
    450   455   460

Glu Lys Gly Glu Ser Val
465   470

<210> 25
<211> 391
<212> PRT
<213> Branchiostoma lanceolatum

<400> 25
Met Ser Ala Asn Thr Thr Val Ser Pro Thr Glu Thr Thr Ala Asn Leu
  1   5   10   15

Thr Ala Asn Ser Thr Glu Ala Ser Val Gly Ser Cys Phe Ala Pro Asn
                20   25   30

Pro Tyr Ser Ala Gly Val Gln Ala Val Leu Gly Leu Ile Thr Val Ile
        35   40   45

Leu Ile Leu Leu Thr Val Ile Gly Asn Val Leu Val Ile Leu Ala Val
    50   55   60

Thr Cys His Arg Lys Met Arg Thr Val Thr Asn Phe Phe Ile Val Ser
65   70   75   80

Leu Ala Cys Ala Asp Leu Ser Val Gly Ile Thr Val Leu Pro Phe Ala
                85   90   95

Ala Thr Asn Asp Ile Leu Gly Tyr Trp Pro Phe Gly Gly Tyr Cys Asp
                100   105   110

Val Trp Val Ser Phe Asp Val Leu Asn Ser Thr Ala Ser Ile Leu Asn
        115   120   125

Leu Val Val Ile Ala Phe Asp Arg Phe Leu Ala Ile Thr Ala Pro Phe
    130   135   140

Thr Tyr His Thr Arg Met Thr Glu Arg Thr Ala Gly Ile Leu Ile Ala
145   150   155   160

Thr Val Trp Gly Ile Ser Leu Val Val Ser Phe Leu Pro Ile Gln Ala
                165   170   175

Gly Trp Tyr Arg Asp Asn Gln Ser Glu Glu Ala Leu Ala Ile Tyr Ser
            180   185   190

Asp Pro Cys Leu Cys Ile Phe Thr Ala Ser Thr Ala Tyr Thr Ile Val
            195   200   205

Ser Ser Leu Ile Ser Phe Tyr Ile Pro Leu Leu Ile Met Leu Val Phe
    210   215   220

Tyr Gly Ile Ile Phe Lys Ala Ala Arg Asp Gln Ala Arg Lys Ile Asn
225   230   235   240

```
Ala Leu Glu Gly Arg Leu Glu Gln Glu Asn Asn Arg Gly Lys Lys Ile
              245             250             255

Ser Leu Ala Lys Glu Lys Lys Ala Ala Lys Thr Leu Gly Ile Ile Met
              260             265             270

Gly Val Phe Ile Leu Cys Trp Leu Pro Phe Phe Val Val Asn Ile Val
              275             280             285

Asn Pro Phe Cys Asp Arg Cys Val Gln Pro Ala Val Phe Ile Ala Leu
              290             295             300

Thr Trp Leu Gly Trp Ile Asn Ser Cys Phe Asn Pro Ile Ile Tyr Ala
305               310             315             320

Phe Asn Lys Glu Phe Arg Lys Val Phe Val Lys Met Ile Cys Cys His
              325             330             335

Lys Cys Arg Gly Val Thr Val Gly Pro Asn His Ala Asp Leu Asn Tyr
              340             345             350

Asp Pro Val Ala Met Arg Leu Lys Lys Arg Gly Glu Asn Ala Asn Gly
              355             360             365

Thr Val Asn Gly Asp Ala Asn Gly Lys Ala Asn Gly Asn Ile Glu Ala
              370             375             380

Gly Glu Gly Thr Ser Ser Ser
385               390
```

```
<210> 26
<211> 36
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthesized
      peptide

<400> 26
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
  1           5               10              15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
              20              25              30

Ile Arg Ser Thr
              35
```

```
<210> 27
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthesized
```

peptide

<400> 27
Gln Arg Lys Pro Gln Leu Leu Gln Val Thr Asn Arg Phe
  1   5   10

<210> 28
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthesized
      peptide

<400> 28
Trp Pro Leu Asn Ser
  1   5

<210> 29
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthesized
      peptide

<400> 29
Asp Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met
  1   5   10   15

Thr Gln Arg Arg
              20

<210> 30
<211> 23
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthesized
      peptide

<400> 30
Gly Gln Ala Ala Phe Asp Glu Arg Asn Ala Leu Cys Ser Met Ile Trp
  1   5   10   15

Gly Ala Ser Pro Ser Tyr Thr
              20

<210> 31
<211> 182
<212> PRT
<213> Artificial Sequence

36

<220>
<223> Description of Artificial Sequence: Synthesized
      peptide

<400> 31
Cys Ala Ala Arg Arg Gln His Ala Leu Leu Tyr Asn Val Lys Arg His
 1   5   10   15

Ser Leu Glu Val Arg Val Lys Asp Cys Val Glu Asn Glu Asp Glu Glu
            20   25   30

Gly Ala Glu Lys Lys Glu Glu Phe Gln Asp Glu Ser Glu Phe Arg Arg
            35   40   45

Gln His Glu Gly Glu Val Lys Ala Lys Glu Gly Arg Met Glu Ala Lys
    50   55   60

Asp Gly Ser Leu Lys Ala Lys Glu Gly Ser Thr Gly Thr Ser Glu Ser
65   70   75   80

Ser Val Glu Ala Gly Ser Glu Glu Val Arg Glu Ser Ser Thr Val Ala
                85   90   95

Ser Asp Gly Ser Met Glu Gly Lys Glu Gly Ser Thr Lys Val Glu Glu
            100   105   110

Asn Ser Met Lys Ala Asp Lys Gly Arg Thr Glu Val Asn Gln Cys Ser
        115   120   125

Ile Asp Leu Gly Glu Asp Asp Met Glu Phe Gly Glu Asp Asp Ile Asn
    130   135   140

Phe Ser Glu Asp Asp Val Glu Ala Val Asn Ile Pro Glu Ser Leu Pro
145   150   155   160

Pro Ser Arg Arg Asn Ser Asn Ser Asn Pro Pro Leu Pro Arg Cys Tyr
                165   170   175

Gln Cys Lys Ala Ala Lys
            180

<210> 32
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthesized
      peptide

<400> 32
Ala Val Leu Ala Val Trp Val Asp Val Glu Thr Gln Val Pro Gln
 1   5   10   15

<210> 33
<211> 55

```
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   Synthesized
      peptide

<400> 33
Tyr Gly Tyr Met His Lys Thr Ile Lys Lys Glu Ile Gln Asp Met Leu
  1           5               10                  15

Lys Lys Phe Phe Cys Lys Glu Lys Pro Pro Lys Glu Asp Ser His Pro
             20              25              30

Asp Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly Lys Ile Val Pro Ser
         35       .       40                  45

Tyr Asp Ser Ala Thr Phe Pro
     50              55


<210> 34
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   HGPRBMY8
      sense primer

<400> 34
gcagagcact cctccactct                                           20


<210> 35
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   HGPRBMY8
      anti-sense primer

<400> 35
agcaggcaat catgacaatc                                           20


<210> 36
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   GPCR84 sense
      primer

<400> 36
gttagcctca cccacctgtt                                           20
```

```
<210> 37
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  GPCR84
      anti-sense primer

<400> 37
cacaatccag gtgccataga                                                    20


<210> 38
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  HGPRBMY8 5'
      primer

<400> 38
gtccccaagc ttgcaccatg acgtccacct gcaccaacag ca                          42


<210> 39
<211> 62
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  HGPRBMY8 3'
      Flag-tag primer

<400> 39
cgggatccta cttgtcgtcg tcgtccttgt agtccatagg aaaagtagca gaatcgtagg 60
aa                                                                      62


<210> 40
<211> 407
<212> PRT
<213> Homo sapiens

<400> 40
Met Ser Leu Asn Ser Ser Leu Ser Cys Arg Lys Glu Leu Ser Asn Leu
  1               5                  10                  15

Thr Glu Glu Glu Gly Gly Glu Gly Gly Val Ile Ile Thr Gln Phe Ile
            20                  25                  30

Ala Ile Ile Val Ile Thr Ile Phe Val Cys Leu Gly Asn Leu Val Ile
          35                  40                  45

Val Val Thr Leu Tyr Lys Lys Ser Tyr Leu Leu Thr Leu Ser Asn Lys
        50                  55                  60
```

39

```
Phe Val Phe Ser Leu Thr Leu Ser Asn Phe Leu Leu Ser Val Leu Val
 65              70              75              80

Leu Pro Phe Val Val Thr Ser Ser Ile Arg Arg Glu Trp Ile Phe Gly
             85              90              95

Val Val Trp Cys Asn Phe Ser Ala Leu Leu Tyr Leu Leu Ile Ser Ser
         100             105             110

Ala Ser Met Leu Thr Leu Gly Val Ile Ala Ile Asp Arg Tyr Tyr Ala
     115             120             125

Val Leu Tyr Pro Met Val Tyr Pro Met Lys Ile Thr Gly Asn Arg Ala
     130             135             140

Val Met Ala Leu Val Tyr Ile Trp Leu His Ser Leu Ile Gly Cys Leu
145             150             155             160

Pro Pro Leu Phe Gly Trp Ser Ser Val Glu Phe Asp Glu Phe Lys Trp
             165             170             175

Met Cys Val Ala Ala Trp His Arg Glu Pro Gly Tyr Thr Ala Phe Trp
         180             185             190

Gln Ile Trp Cys Ala Leu Phe Pro Phe Leu Val Met Leu Val Cys Tyr
         195             200             205

Gly Phe Ile Phe Arg Val Ala Arg Val Lys Ala Arg Lys Val His Cys
     210             215             220

Gly Thr Val Val Ile Val Glu Glu Asp Ala Gln Arg Thr Gly Arg Lys
225             230             235             240

Asn Ser Ser Thr Ser Thr Ser Ser Ser Gly Ser Arg Arg Asn Ala Phe
             245             250             255

Gln Gly Val Val Tyr Ser Ala Asn Gln Cys Lys Ala Leu Ile Thr Ile
         260             265             270

Leu Val Val Leu Gly Ala Phe Met Val Thr Trp Gly Pro Tyr Met Val
     275             280             285

Val Ile Ala Ser Glu Ala Leu Trp Gly Lys Ser Ser Val Ser Pro Ser
     290             295             300

Leu Glu Thr Trp Ala Thr Trp Leu Ser Phe Ala Ser Ala Val Cys His
305             310             315             320

Pro Leu Ile Tyr Gly Leu Trp Asn Lys Thr Val Arg Lys Glu Leu Leu
             325             330             335

Gly Met Cys Phe Gly Asp Arg Tyr Tyr Arg Glu Pro Phe Val Gln Arg
         340             345             350

Gln Arg Thr Ser Arg Leu Phe Ser Ile Ser Asn Arg Ile Thr Asp Leu
         355             360             365

Gly Leu Ser Pro His Leu Thr Ala Leu Met Ala Gly Gly Gln Pro Leu
     370             375             380
```

40

Gly His Ser Ser Ser Thr Gly Asp Thr Gly Phe Ser Cys Ser Gln Asp
385   390   395   400

Ser Gly Asn Leu Arg Ala Leu
              405

<210> 41
<211> 448
<212> PRT
<213> Homo sapiens

<400> 41
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
 1   5   10   15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
              20   25   30

Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
          35   40   45

Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
      50   55   60

Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
 65   70   75   80

Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
              85   90   95

Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
          100   105   110

Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Val Val Ser Val Asp
          115   120   125

Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
      130   135   140

Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145   150   155   160

Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp
              165   170   175

Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
          180   185   190

Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
          195   200   205

Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
      210   215   220

Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225   230   235   240

```
Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
              245               250               255

Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
          260               265               270

Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
          275               280               285

Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
      290               295               300

Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305               310               315               320

Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys
              325               330               335

Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Asp
          340               345               350

Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
          355               360               365

Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
    370               375               380

Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Lys Lys Val
385               390               395               400

Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
              405               410               415

Phe Leu Ala Val Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly
          420               425               430

Gly Thr Glu Gly Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
          435               440               445
```

<210> 42
<211> 448
<212> PRT
<213> Homo sapiens

<400> 42
```
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
  1               5               10               15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
          20               25               30

Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
          35               40               45

Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
```

          50   55   60

Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
65   70   75   80

Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
              85   90   95

Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
          100   105   110

Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Val Val Ser Val Asp
          115   120   125

Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
      130   135   140

Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145   150   155   160

Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp
              165   170   175

Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
          180   185   190

Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
      195   200   205

Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
      210   215   220

Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225   230   235   240

Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
              245   250   255

Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
          260   265   270

Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
          275   280   285

Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
          290   295   300

Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305   310   315   320

Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys
              325   330   335

Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Asp
          340   345   350

Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
          355   360   365

```
Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
    370             375             380

Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Lys Lys Val
385             390             395                 400

Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
            405             410             415

Phe Leu Ala Val Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly
            420             425             430

Gly Thr Glu Gly Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
            435             440             445
```

```
<210> 43
<211> 448
<212> PRT
<213> Homo sapiens

<400> 43
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
  1             5               10                  15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
            20              25                  30

Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
        35              40                  45

Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
    50              55                  60

Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
65              70              75                  80

Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
            85              90                  95

Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
            100             105             110

Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Leu Val Ser Val Asp
        115             120             125

Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
    130             135             140

Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145             150             155                 160

Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp
            165             170             175

Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
```

44

```
              180                    185                    190
Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
         195                    200                    205

Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
         210                    215                    220

Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225                    230                    235                    240

Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
                   245                    250                    255

Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
                   260                    265                    270

Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
              275                    280                    285

Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
         290                    295                    300

Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305                    310                    315                    320

Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys
                   325                    330                    335

Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Gly
              340                    345                    350

Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
         355                    360                    365

Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
370                    375                    380

Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Ala Lys Val
385                    390                    395                    400

Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
                   405                    410                    415

Phe Leu Ala Val Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly
         420                    425                    430

Gly Thr Glu Gly Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
         435                    440                    445
```

```
<210> 44
<211> 1659
<212> DNA
<213> Homo sapiens
```

45

<400> 44
gcctgcaacc tgtcycacgc cctctggctg ttgccatgac gtccacctgc accaacagca 60
cgcgcgagag taacagcagc cacacgtgca tgcccctctc caaaatgccc atcagcctgg 120
cccacggcat catccgctca accgtgctgg ttatcttcct cgccgcctct ttcgtcggca 180
acatagtgct ggcgctagtg ttgcagcgca agccgcagct gctgcaggtg accaaccgtt 240
ttatctttaa cctcctcgtc accgacctgc tgcagatttc gctcgtggcc ccctgggtgg 300
tggccacctc tgtgcctctc ttctggcccc tcaacagcca cttctgcacg gccctggtta 360
gcctcaccca cctgttcgcc ttcgccagcg tcaacaccat tgtcttggtg tcagtggatc 420
gctacttgtc catcatccac cctctctcct accgtccaa gatgacccag cgccgcggtt 480
acctgctcct ctatggcacc tggattgtgg ccatcctgca gagcactcct ccactctacg 540
gctggggcca ggctgccttt gatgagcgca atgctctctg ctccatgatc tggggggcca 600
gccccagcta cactattctc agcgtggtgt ccttcatcgt cattccactg attgtcatga 660
ttgcctgcta ctccgtggtg ttctgtgcag cccggaggca gcatgctctg ctgtacaatg 720
tcaagagaca cagcttggaa gtgcgagtca aggactgtgt ggagaatgag gatgaagagg 780
gagcagagaa gaaggaggag ttccaggatg agagtgagtt tcgccgccag catgaaggtg 840
aggtcaaggc caaggagggc agaatggaag ccaaggacgg cagcctgaag gccaaggaag 900
gaagcacggg gaccagtgag agtagtgtag aggccagggg cagcgaggag gtcagagaga 960
gcagcacggt ggccagcgac ggcagcatgg agggtaagga aggcagcacc aaagttgagg 1020
agaacagcat gaaggcagac aagggtcgca cagaggtcaa ccagtgcagc attgacttgg 1080
gtgaagatgg catggagttt ggtgaagacg acatcaattt cagtgaggat gacgtcgagg 1140
cagtgaacat cccggagagc ctcccaccca gtcgtcgtaa cagcaacagc aaccctcctc 1200
tgcccaggtg ctaccagtgc aaagctgcta aagtgatctt catcatcatt ttctcctatg 1260
tgctatccct ggggccctac tgctttttag cagtcctggc cgtgtgggtg gatgtcgaaa 1320
cccaggtacc ccagtgggtg atcaccataa tcatctggct tttcttcctg cagtgctgca 1380
tccaccccta tgtctatggc tacatgcaca agaccattaa gaaggaaatc caggacatgc 1440
tgaagaagtt cttctgcaag gaaaagcccc gaaagaaga tagccaccca gacctgcccg 1500
gaacagaggg tgggactgaa ggcaagattg tcccttccta cgattctgct acttttcctt 1560
gaagttagtt ctaaggcaaa ccttgaaaat cagtccttca gccacagcta tttagagctt 1620
taaaactacc aggttcaatc actggttatg ctttctgtg 1659

<210> 45
<211> 1527
<212> DNA
<213> Homo sapiens

<400> 45
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac gtgcatgccc 60
ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca gcgcaagccg 180
cagctgctgc aggtgaccaa ccgttttatc tttaacctcc tcgtcaccga cctgctgcag 240
atttcgctcg tggcccctg ggtggtggcc acctctgtgc ctctcttctg gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc cagcgtcaac 360
accattgtcg tggtgtcagt ggatcgctac ttgtccatca tccaccctct ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg gcacctggat tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt ggtgtccttc 600

```
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct tggaagtgcg agtcaaggac 720
tgtgtggaga atgaggatga agagggagca gagaagaagg aggagttcca ggatgagagt 780
gagtttcgcc gccagcatga aggtgaggtc aaggccaagg agggcagaat ggaagccaag 840
gacggcagcc tgaaggccaa ggaaggaagc acggggacca gtgagagtag tgtagaggcc 900
aggggcagcg aggaggtcag agagagcagc acggtggcca gcgacggcag catggagggt 960
aaggaaggca gcaccaaagt tgaggagaac agcatgaagg cagacaaggg tcgcacagag
1020
gtcaaccagt gcagcattga cttgggtgaa gatgacatgg agtttggtga agacgacatc
1080
aatttcagtg aggatgacgt cgaggcagtg aacatcccgg agagcctccc acccagtcgt
1140
cgtaacagca acagcaaccc tcctctgccc aggtgctacc agtgcaaagc tgctaaagtg
1200
atcttcatca tcattttctc ctatgtgcta tccctggggc cctactgctt tttagcagtc
1260
ctggccgtgt gggtggatgt cgaaacccag gtaccccagt gggtgatcac cataatcatc
1320
tggctttttct tcctgcagtg ctgcatccac ccctatgtct atggctacat gcacaagacc
1380
attaagaagg aaatccagga catgctgaag aagttcttct gcaaggaaaa gccccccgaaa
1440
gaagatagcc acccagacct gcccggaaca gagggtggga ctgaaggcaa gattgtccct
1500
tcctacgatt ctgctacttt tccttga
1527
```

```
<210> 46
<211> 1527
<212> DNA
<213> Homo sapiens

<400> 46
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac gtgcatgccc 60
ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca gcgcaagccg 180
cagctgctgc aggtgaccaa ccgtttatc tttaacctcc tcgtcaccga cctgctgcag 240
atttcgctcg tggcccctctg ggtggtggcc acctctgtgc ctctcttctg gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc cagcgtcaac 360
accattgtcg tggtgtcagt ggatcgctac ttgtccatca tccaccctct ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg gcacctggat tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt ggtgtccttc 600
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct tggaagtgcg agtcaaggac 720
tgtgtggaga atgaggatga agagggagca gagaagaagg aggagttcca ggatgagagt 780
gagtttcgcc gccagcatga aggtgaggtc aaggccaagg agggcagaat ggaagccaag 840
gacggcagcc tgaaggccaa ggaaggaagc acggggacca gtgagagtag tgtagaggcc 900
aggggcagcg aggaggtcag agagagcagc acggtggcca gcgacggcag catggagggt 960
aaggaaggca gcaccaaagt tgaggagaac agcatgaagg cagacaaggg tcgcacagag
1020
gtcaaccagt gcagcattga cttgggtgaa gatgacatgg agtttggtga agacgacatc
1080
aatttcagtg aggatgacgt cgaggcagtg aacatcccgg agagcctccc acccagtcgt
1140
cgtaacagca acagcaaccc tcctctgccc aggtgctacc agtgcaaagc taagaaagtg
1200
```

```
atcttcatca tcattttctc ctatgtgcta tccctggggc cctactgctt tttagcagtc
1260
ctggccgtgt gggtggatgt cgaaacccag gtaccccagt gggtgatcac cataatcatc
1320
tggcttttct tcctgcagtg ctgcatccac ccctatgtct atggctacat gcacaagacc
1380
attaagaagg aaatccagga catgctgaag aagttcttct gcaaggaaaa gccccgaaa
1440
gaagatagcc acccagacct gcccggaaca gagggtggga ctgaaggcaa gattgtccct
1500
tcctacgatt ctgctacttt tccttga
1527
```

```
<210> 47
<211> 1580
<212> DNA
<213> Homo sapiens

<400> 47
gcaacctgtc tcacgccctc tggctgttgc catgacgtcc acctgcacca acagcacgcg 60
cgagagtaac agcagccaca cgtgcatgcc cctctccaaa atgcccatca gcctggccca 120
cggcatcatc cgctcaaccg tgctggttat cttcctcgcc gcctctttcg tcggcaacat 180
agtgctggcg ctagtgttgc agcgcaagcc gcagctgctg caggtgacca accgttttat 240
ctttaacctc ctcgtcaccg acctgctgca gatttcgctc gtggccccct gggtggtggc 300
cacctctgtg cctctcttct ggcccctcaa cagccacttc tgcacggccc tggttagcct 360
cacccacctg ttcgccttcg ccagcgtcaa caccattgtc ttggtgtcag tggatcgcta 420
cttgtccatc atccaccctc tctcctaccc gtccaagatg acccagcgcc gcggttacct 480
gctcctctat ggcacctgga ttgtggccat cctgcagagc actcctccac tctacggctg 540
gggccaggct gcctttgatg agcgcaatgc tctctgctcc atgatctggg gggccagccc 600
cagctacact attctcagcg tggtgtcctt catcgtcatt ccactgattg tcatgattgc 660
ctgctactcc gtggtgttct gtgcagcccg gaggcagcat gctctgctgt acaatgtcaa 720
gagacacagc ttggaagtgc gagtcaagga ctgtgtggag aatgaggatg aagagggagc 780
agagaagaag gaggagttcc aggatgagag tgagtttcgc cgccagcatg aaggtgaggt 840
caaggccaag gagggcagaa tggaagccaa ggacggcagc ctgaaggcca aggaaggaag 900
cacgggacc agtgagagta gtgtagaggc caggggcagc gaggaggtca gagagagcag 960
cacggtggcc agcgacggca gcatggaggg taaggaaggc agcaccaaag ttgaggagaa
1020
cagcatgaag gcagacaagg gtcgcacaga ggtcaaccag tgcagcattg acttgggtga
1080
agatgacatg gagtttggtg aagacgacat caatttcagt gaggatgacg tcgaggcagt
1140
gaacatcccg gagagcctcc cacccagtcg tcgtaacagc aacagcaacc ctcctctgcc
1200
caggtgctac cagtgcaaag ctgctaaagt gatcttcatc atcattttct cctatgtgct
1260
atccctgggg ccctactgct ttttagcagt cctggccgtg tgggtggatg tcgaaaccca
1320
ggtaccccag tgggtgatca ccataatcat ctggcttttc ttcctgcagt gctgcatcca
1380
cccctatgtc tatggctaca tgcacaagac cattaagaag gaaatccagg acatgctgaa
1440
gaagttcttc tgcaaggaaa gcccccgaa agaagatagc cacccagacc tgcccggaac
1500
agagggtggg actgaaggca agattgtccc ttcctacgat tctgctactt ttccttgaag
1560
ttagttctaa ggcaaacctt
1580
```

<210> 48
<211> 1527
<212> DNA
<213> Homo sapiens

<220>
<223> N=A+T+G+C

<400> 48
atgacgtcca cctgcaccaa cagcacgcgc gagagtaaca gcagccacac gtgcatgccc 60
ctctccaaaa tgcccatcag cctggcccac ggcatcatcc gctcaaccgt gctggttatc 120
ttcctcgccg cctctttcgt cggcaacata gtgctggcgc tagtgttgca gcgcaagccg 180
cagctgctgc aggtgaccaa ccgttttatc tttaacctcc tcgtcaccga cctgctgcag 240
atttcgctcg tggccccctg ggtggtggcc acctctgtgc ctctcttctg gcccctcaac 300
agccacttct gcacggccct ggttagcctc acccacctgt tcgccttcgc cagcgtcaac 360
accattgtcn tggtgtcagt ggatcgctac ttgtccatca tccaccctct ctcctacccg 420
tccaagatga cccagcgccg cggttacctg ctcctctatg gcacctggat tgtggccatc 480
ctgcagagca ctcctccact ctacggctgg ggccaggctg cctttgatga gcgcaatgct 540
ctctgctcca tgatctgggg ggccagcccc agctacacta ttctcagcgt ggtgtccttc 600
atcgtcattc cactgattgt catgattgcc tgctactccg tggtgttctg tgcagcccgg 660
aggcagcatg ctctgctgta caatgtcaag agacacagct tggaagtgcg agtcaaggac 720
tgtgtggaga tgaggatga agagggagca gagaagaagg aggagttcca ggatgagagt 780
gagtttcgcc gccagcatga aggtgaggtc aaggccaagg agggcagaat ggaagccaag 840
gacggcagcc tgaaggccaa ggaaggaagc acggggacca gtgagagtag tgtagaggcc 900
aggggcagcg aggaggtcag agagagcagc acggtggcca gcacggcag catggagggt 960
aaggaaggca gcaccaaagt tgaggagaac agcatgaagg cagacaaggg tcgcacagag
1020
gtcaaccagt gcagcattga cttgggtgaa gatgncatgg agtttggtga agacgacatc
1080
aatttcagtg aggatgacgt cgaggcagtg aacatcccgg agagcctccc acccagtcgt
1140
cgtaacagca acagcaaccc tcctctgccc aggtgctacc agtgcaaagc tnnnaaagtg
1200
atcttcatca tcattttctc ctatgtgcta tccctggggc cctactgctt tttagcagtc
1260
ctggccgtgt gggtggatgt cgaaacccag gtaccccagt gggtgatcac cataatcatc
1320
tggctttttct tcctgcagtg ctgcatccac ccctatgtct atggctacat gcacaagacc
1380
attaagaagg aaatccagga catgctgaag aagttcttct gcaaggaaaa gccccccgaaa
1440
gaagatagcc acccagacct gcccggaaca gagggtggga ctgaaggcaa gattgtccct
1500
tcctacgatt ctgctacttt tccttga
1527

<210> 49
<211> 508
<212> PRT
<213> Homo sapiens

<220>
<223> Xaa=Unknown, modified, or any amino acid

<400> 49
Met Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
  1   5   10   15

Thr Cys Met Pro Leu Ser Lys Met Pro Ile Ser Leu Ala His Gly Ile
          20   25   30

Ile Arg Ser Thr Val Leu Val Ile Phe Leu Ala Ala Ser Phe Val Gly
          35   40   45

Asn Ile Val Leu Ala Leu Val Leu Gln Arg Lys Pro Gln Leu Leu Gln
    50   55   60

Val Thr Asn Arg Phe Ile Phe Asn Leu Leu Val Thr Asp Leu Leu Gln
65   70   75   80

Ile Ser Leu Val Ala Pro Trp Val Val Ala Thr Ser Val Pro Leu Phe
                85   90   95

Trp Pro Leu Asn Ser His Phe Cys Thr Ala Leu Val Ser Leu Thr His
              100   105   110

Leu Phe Ala Phe Ala Ser Val Asn Thr Ile Val Xaa Val Ser Val Asp
          115   120   125

Arg Tyr Leu Ser Ile Ile His Pro Leu Ser Tyr Pro Ser Lys Met Thr
    130   135   140

Gln Arg Arg Gly Tyr Leu Leu Leu Tyr Gly Thr Trp Ile Val Ala Ile
145   150   155   160

Leu Gln Ser Thr Pro Pro Leu Tyr Gly Trp Gly Gln Ala Ala Phe Asp
              165   170   175

Glu Arg Asn Ala Leu Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr
              180   185   190

Thr Ile Leu Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met
          195   200   205

Ile Ala Cys Tyr Ser Val Val Phe Cys Ala Ala Arg Arg Gln His Ala
    210   215   220

Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val Lys Asp
225   230   235   240

Cys Val Glu Asn Glu Asp Glu Glu Gly Ala Glu Lys Lys Glu Glu Phe
              245   250   255

Gln Asp Glu Ser Glu Phe Arg Arg Gln His Glu Gly Glu Val Lys Ala
              260   265   270

Lys Glu Gly Arg Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu
          275   280   285

Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
    290   295   300

Glu Val Arg Glu Ser Ser Thr Val Ala Ser Asp Gly Ser Met Glu Gly
305   310   315   320

Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys

```
             325                    330                    335
```

Gly Arg Thr Glu Val Asn Gln Cys Ser Ile Asp Leu Gly Glu Asp Xaa
          340   345   350

Met Glu Phe Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu
          355   360   365

Ala Val Asn Ile Pro Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn
    370   375   380

Ser Asn Pro Pro Leu Pro Arg Cys Tyr Gln Cys Lys Ala Xaa Lys Val
385   390   395   400

Ile Phe Ile Ile Ile Phe Ser Tyr Val Leu Ser Leu Gly Pro Tyr Cys
              405   410   415

Phe Leu Ala Val Leu Ala Val Trp Val Asp Val Glu Thr Gln Val Pro
          420   425   430

Gln Trp Val Ile Thr Ile Ile Ile Trp Leu Phe Phe Leu Gln Cys Cys
          435   440   445

Ile His Pro Tyr Val Tyr Gly Tyr Met His Lys Thr Ile Lys Lys Glu
    450   455   460

Ile Gln Asp Met Leu Lys Lys Phe Phe Cys Lys Glu Lys Pro Pro Lys
465   470   475   480

Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly
              485   490   495

Lys Ile Val Pro Ser Tyr Asp Ser Ala Thr Phe Pro
          500   505

```
<210> 50
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  SNP

<400> 50
caccattgtc ttggtgtcag t                                            21


<210> 51
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  SNP

<400> 51
caccattgtc gtggtgtcag t                                            21
```

```
<210> 52
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   SNP

<400> 52
ggtgaagatg acatggagtt t                        ..                21


<210> 53
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   SNP

<400> 53
ggtgaagatg gcatggagtt t                                          21


<210> 54
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   SNP

<400> 54
gtgcaaagct gctaaagtga t                                          21


<210> 55
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   SNP

<400> 55
gtgcaaagct actaaagtga t                                          21


<210> 56
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   SNP

<400> 56
tgcaaagctg ctaaagtgat c                                          21
```

```
<210> 57
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  SNP

<400> 57
tgcaaagctg ataaagtgat c                                              21


<210> 58
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  SNP

<400> 58
gcaaagctgc taaagtgatc t                                              21


<210> 59
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  SNP

<400> 59
gcaaagctgc gaaagtgatc t                                              21


<210> 60
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  GAPDH F3
      Forward primer

<400> 60
agccgagcca catcgct                                                   17


<210> 61
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  GAPDH R1
      Reverse primer
```

<400> 61
gtgaccaggc gcccaatac   19

<210> 62
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: GAPDH-PVIC
      Taqman(R) Probe

<400> 62
caaatccgtt gactccgacc ttcacctt   28

<210> 63
<211> 99
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligo 1;
      N=A+G+C+T and B=C+G+T

<400> 63
cgaagcgtaa gggcccagcc ggccnnbnnb nnbnnbnnbn nbnnbnnbnn bnnbnnbnnb 60
nnbnnbnnbn nbnnbnnbnn bnnbccgggt ccgggcggc   99

<210> 64
<211> 95
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligo 2;
      N=A+G+C+T and V=C+A+G

<400> 64
aaaaggaaaa aagcggccgc vnnvnnvnnv nnvnnvnnvn nvnnvnnvnn vnnvnnvnnv 60
nnvnnvnnvn nvnnvnnvnn gccgcccgga cccgg   95

<210> 65
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 65
Pro Gly Pro Gly Gly
  1   5

```
<210> 66
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 66
Gly Asp Phe Trp Tyr Glu Ala Cys Glu Ser Ser Cys Ala Phe Trp
 1           5               10                  15


<210> 67
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 67
Leu Glu Trp Gly Ser Asp Val Phe Tyr Asp Val Tyr Asp Cys Cys
 1           5               10                  15


<210> 68
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 68
Cys Leu Arg Ser Gly Thr Gly Cys Ala Phe Gln Leu Tyr Arg Phe
 1           5               10                  15


<210> 69
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 69
Asn Asn Phe Pro Cys Leu Arg Ser Gly Arg Asn Cys Asp Ala Gly
 1           5               10                  15


<210> 70
<211> 15
```

```
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 70
Arg Ile Val Pro Asn Gly Tyr Phe Asn Val His Gly Arg Ser Leu
 1               5                  10                  15


<210> 71
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic
      polypeptide

<400> 71
Arg Ile Asp Ser Cys Ala Lys Tyr Phe Leu Arg Ser Cys Asp
 1               5                  10


<210> 72
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic 5'
      primer

<400> 72
gcagcagcgg ccgcaccgtg ctggttatct tcctcgccg                         39


<210> 73
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic 3'
      primer

<400> 73
gcagcagtcg acaggaaaag tagcagaatc gtagg                             35


<210> 74
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:  Synthetic 5'
```

primer

<400> 74
gcagcagcgg ccgcatgacg tccacctgca ccaacagc   38

<210> 75
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic 3'
      primer

<400> 75
gcagcagtcg acatagacat aggggtggat gcagcac   37

<210> 76
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 76
Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser
 1   5   10

<210> 77
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 77
Gln Leu Leu Gln Val Thr Asn Arg Phe Ile Phe Asn Leu
 1   5   10

<210> 78
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 78
Tyr Pro Ser Lys Met Thr Gln Arg Arg Gly Tyr Leu Leu
 1   5   10

<210> 79
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 79
Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu Gly Ser
  1   5   10

<210> 80
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 80
Glu Gly Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser
  1   5   10

<210> 81
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 81
Lys Val Glu Glu Asn Ser Met Lys Ala Asp Lys Gly Arg
  1   5   10

<210> 82
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 82
Glu Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn Ser Asn
  1   5   10

<210> 83

<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 83
Gly Tyr Met His Lys Thr Ile Lys Lys Glu Ile Gln Asp
  1   5   10

<210> 84
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 84
Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser His
  1   5   10

<210> 85
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 85
Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly Ser Glu
  1   5   10

<210> 86
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 86
Gly Lys Glu Gly Ser Thr Lys Val Glu Glu Asn Ser Met Lys
  1   5   10

<210> 87
<211> 14
<212> PRT
<213> Artificial Sequence

59

```
<220>
<223> Description of Artificial Sequence:   Synthetic
      polypeptide

<400> 87
Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu Ala Val Asn
  1               5                  10


<210> 88
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   Synthetic
      polypeptide

<400> 88
Pro Pro Lys Glu Asp Ser His Pro Asp Leu Pro Gly Thr Glu
  1               5                  10


<210> 89
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   Synthetic
      polypeptide

<400> 89
Leu Leu Tyr Asn Val Lys Arg His Ser Leu Glu Val Arg Val
  1               5                  10


<210> 90
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   Synthetic
      polypeptide

<400> 90
Ser Leu Pro Pro Ser Arg Arg Asn Ser Asn Ser Asn Pro Pro
  1               5                  10


<210> 91
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:   Synthetic
```

polypeptide

<400> 91
Thr Ser Thr Cys Thr Asn Ser Thr Arg Glu Ser Asn Ser Ser
 1   5   10

<210> 92
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 92
Ser Thr Arg Glu Ser Asn Ser Ser His Thr Cys Met Pro Leu
 1   5   10

<210> 93
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 93
Gly Glu Asp Asp Ile Asn Phe Ser Glu Asp Asp Val Glu Ala
 1   5   10

<210> 94
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 94
Ile Ser Leu Ala His Gly Ile Ile Arg Ser Thr Val Leu Val Ile Phe
 1   5   10   15

<210> 95
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 95

```
Cys Ser Met Ile Trp Gly Ala Ser Pro Ser Tyr Thr Ile Leu Ser Val
1               5                   10                  15
```


<210> 96
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 96
```
Met Glu Ala Lys Asp Gly Ser Leu Lys Ala Lys Glu Gly Ser Thr Gly
1               5                   10                  15
```


<210> 97
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 97
```
Leu Lys Ala Lys Glu Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu
1               5                   10                  15
```


<210> 98
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 98
```
Lys Glu Gly Ser Thr Gly Thr Ser Glu Ser Ser Val Glu Ala Arg Gly
1               5                   10                  15
```


<210> 99
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 99
```
Thr Val Ala Ser Asp Gly Ser Met Glu Gly Lys Glu Gly Ser Thr Lys
1               5                   10                  15
```

<210> 100
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 100
His Pro Asp Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly Lys Ile Val
 1   5   10   ,   15

<210> 101
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 101
Leu Pro Gly Thr Glu Gly Gly Thr Glu Gly Lys Ile Val Pro Ser Tyr
 1   5   10   15

<210> 102
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 102
Ser Val Val Ser Phe Ile Val Ile Pro Leu Ile Val Met Ile Ala Cys
 1   5   10   15   .

Tyr Ser Val Val Phe
          20

ATCC

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF
THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT ISSUED PURSUANT TO RULE 7.3
AND VIABILITY STATEMENT ISSUED PURSUANT TO RULE 10.2

To: (Name and Address of Depositor or Attorney)

       Bristol-Myers Squibb Company
       Attn: John Feder
       P.O. Box 5400
       Princeton, NJ 08543

Deposited on Behalf of: Bristol-Myers Squibb Company

Identification Reference by Depositor:	**Patent Deposit Designation**
Human cDNA inserts cloned into vector pSPORT; gene names are-	PTA-2966
HGPRBMY8, HGPRBMY23, BMY-HPP5, HGPRBMY7, CGR1,	
K+betaM2, K+alphaM1 (FL): BMS Group B	

The deposit was accompanied by: __ a scientific description __ a proposed taxonomic description indicated above.

The deposit was received January 24, 2001 by this International Depository Authority and has been accepted.

AT YOUR REQUEST: X   We will inform you of requests for the strain for 30 years.

The strain will be made available if a patent office signatory to the Budapest Treaty certifies one's right to receive, or if a U.S. Patent is issued citing the strain, and ATCC is instructed by the United States Patent & Trademark Office or the depositor to release said strain.

If the culture should die or be destroyed during the effective term of the deposit, it shall be your responsibility to replace it with living culture of the same.

The strain will be maintained for a period of at least 30 years from date of deposit, or five years after the most recent request for a sample, whichever is longer. The United States and many other countries are signatory to the Budapest Treaty.

The viability of the culture cited above was tested January 31, 2001. On that date, the culture was viable.

International Depository Authority: American Type Culture Collection, Manassas, VA 20110-2209 USA.

Signature of person having authority to represent ATCC:

Tanya Nunnally, Patent Specialist, Patent Depository   Date: February 5, 2001

cc: Stephen Damico

(54) Title: A HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMY8, EXPRESSED HIGHLY IN BRAIN

(57) Abstract: The present invention describes a newly discovered human G-protein coupled receptor and its encoding polynucleotide. Also described are expression vectors, host cells, agonists, antagonists, antisense molecules, and antibodies associated with the polynucleotide and/or polypeptide of the present invention. In addition, methods for treating, diagnosing, preventing, and screening for disorders associated with aberrant cell growth, neurological conditions, and diseases or disorders related to the brain are illustrated.

International Application No

PCT/US 01/43909

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   C12N15/12   C07K14/705   C12N1/21   C12N5/10   C12N15/62
        A61K38/17   C07K16/28   G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   C12N   C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

SEQUENCE SEARCH, EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 05274 A (WILSON SHELAGH ;SMITHKLINE BEECHAM PLC (GB); SMITHKLINE BEECHAM CO) 4 February 1999 (1999-02-04) the whole document	1-22
E	GB 2 367 295 A (SMITHKLINE BEECHAM PLC ;SMITHKLINE BEECHAM CORP (US)) 3 April 2002 (2002-04-03) the whole document SEQ ID NO:2	4,6,7, 10,11, 13,15, 19-22
E	EP 1 243 648 A (MITSUBISHI PHARMA CORP) 25 September 2002 (2002-09-25) the whole document SEQ ID NO:21	1-4, 6-12,14, 15,19-22

☐ Further documents are listed in the continuation of box C.   ☒ Patent family members are listed in annex.

* Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance
'E' earlier document but published on or after the international filing date
'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
'O' document referring to an oral disclosure, use, exhibition or other means
'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 November 2002	20/11/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Sprinks, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Box I   Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

   Although claims 11 and 15 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II   Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (1)) (July 1998)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9905274	A	04-02-1999	WO	9905274 A1	04-02-1999
GB 2367295	A	03-04-2002	NONE		
EP 1243648	A	25-09-2002	AU	2230301 A	09-07-2001
			EP	1243648 A1	25-09-2002
			AU	2230401 A	09-07-2001
			WO	0148188 A1	05-07-2001
			WO	0148189 A1	05-07-2001
			AU	4467301 A	08-10-2001
			WO	0173023 A1	04-10-2001

THIS PAGE BLANK (USPTO)